

Science

4 June 2010 | \$10



AAAS



Ken-ichi Arai is professor emeritus at the University of Tokyo, founding president of the Asia-Pacific International Molecular Biology Network, and president and chief executive officer of SBI Biotech.

Japanese Science in a Global World

THE BEGINNING OF 2010 BROUGHT GOOD ECONOMIC NEWS TO JAPAN, WHICH HAS SEEN ONLY lackluster growth for the past two decades. Although the country's debt load and deflation remain serious problems, the world's second largest economy reportedly grew in the last quarter of 2009. But that does not mean that Japan is on a route to long-term recovery. Although the science budget for fiscal year 2011 was not severely cut, a worrisome sign was the government's attempts to freeze investment in Japan's science infrastructure (for example, supercomputing) and reduce spending on earth sciences, cosmology, and other fields. These were avoided through the protests of prominent Japanese scientists. The lack of political interest in bolstering investment in science and technology indicates misguided thinking. There is therefore a growing awareness in the Japanese research community that scientists need to become more involved in formulating the country's science policy and in guiding young scientists into international networks that will support a successful global economy for Japan.

After World War II, Japan forged a robust economy by focusing on manufacturing industries. But a sharp decrease in global demand for goods, the collapse of an asset bubble, and a failure to quickly adapt its traditional organizational structure in science, technology, and industry drove Japan into an economic crash in the early 1990s, leaving the country incapable of quickly responding through entrepreneurship and new ventures. In response, the Japanese government led by the Liberal Democratic Party instituted reforms in science and technology to encourage innovation. The goals included increasing the number of Ph.D. students and postdoctoral fellows and promoting goal-oriented projects to create new industries.

As a result, the number of graduate students and postdocs and the number of publications by Japanese scientists in prestigious journals increased. However, there has been a downside. The shift to a "big science" view with a "top-down" goal-oriented style has concentrated funding in fewer but bigger projects, thus supporting fewer researchers, rather than funding more individual scientists and small, more focused research endeavors. This has created uncertainty for young researchers today about their career paths in a top-down research environment. This is reflected in a recent decrease in Japanese Ph.D.s going abroad to expand their professional and educational development and to create collaborative relationships. From the 1960s through the 1990s, many Japanese researchers went to North America and Europe for postdoctoral training. Today, young Japanese researchers who face job opportunity challenges may fear not finding a job upon returning home after time abroad. Whatever the precise reasons, the result is an inward-looking attitude of young scientists, which works against establishing strong international scientific networks. In contrast, other Asian countries are energetically nurturing such ties, particularly China, India, Korea, Malaysia, Singapore, and Thailand.

One of the reform goals of the previous Japanese government, which has not been achieved, was to provide independence to young investigators through an open research system that supports international paths for career development without borders (of nationality, gender, or age). Pursuing this goal in the context of a bottom-up open platform for science and technology that frees scientists to pursue their research interests would create career paths that are interchangeable with those in other countries, particularly other Asian nations. Such a flexible research infrastructure would allow for shared opportunities among many Asian countries and may help them tackle common problems in health care, food, energy, and the environment.

After the Democratic Party of Japan was victorious in last year's elections, Prime Minister Yukio Hatoyama's science policy was not initially clear. Today, the good news is that the new government supports genomic and stem cell research, biotech ventures, and green technology programs. But if the economic recovery stalls, science may face severe budget cuts. Scientists must communicate much better with policy-makers about how to direct investments in ways that will put Japan on a path toward sustained growth, including support for a generation of globally integrated and innovative scientists.

— Ken-ichi Arai

10.1126/science.1192830



GULF OIL DISASTER

Louisiana Begins Controversial Engineering to Ward Off Oil Spill

Regardless of when BP finally manages to stop its undersea gusher from the *Deepwater Horizon*, a massive slick will likely remain in the Gulf of Mexico for some time. With public officials desperate for action, the magnitude of the disaster—the largest oil spill in U.S. history—has inspired an unprecedented and untested idea for combating it: an extensive sand trap. But scientists are dubious about the project's chance of success and say it could even jeopardize long-term restoration of Louisiana's wetlands, which have been disappearing for decades.

In mid-May, Louisiana Governor Bobby Jindal—frustrated at what he saw as an inadequate federal response to the spill—proposed a 2-meter-tall sand berm to protect 160 kilometers of his state's coast from oil. The project would require dredging an estimated 68 million cubic meters of sand and cost at least \$350 million—perhaps three times that figure. “I was stunned,” says



First cut. Dredges like this will collect sand, which will be shipped and then dropped in front of barrier islands.

Joseph Kelley, a coastal geophysicist at the University of Maine, Orono. “This is a big proposal and not well thought out.”

Last week, the Army Corps of Engineers, after examining the proposal and getting input from experts at other agencies, granted Louisiana an emergency permit to build two portions of the proposed berm. But scientists at those agencies and at universities question how effective or durable the berm would be, and some say the corps' analysis was too hasty. “You're spending a couple hundred million dollars and crossing your fingers,” says Robert Young, a coastal

geologist at Western Carolina University in Cullowhee, North Carolina.

The rationale for the berm is fairly simple. Oil is much easier to remove from a sandy beach than from a vegetation-rich wetland, where cleaning would probably cause additional harm to the fragile ecosystem. But Louisiana's sandy barrier islands have been greatly damaged by storms in recent years. In some places, the Chandeleur Islands have eroded into shallow shoals over which oil could flow unimpeded toward sensitive wetlands. In addition to catching oil, the massive berms would shunt oily water toward tidal inlets, making it more efficient for booms and boats to collect it.

But soon after the state applied for the emergency permit on 11 May, federal agencies and academic scientists raised objections. “My concern is if we rush into this and don't have any idea about what the impacts might be,” says Gregory Stone of Louisiana State University (LSU) in Baton Rouge. The dredging might exacerbate erosion of the islands, for example, or further disturb breeding birds or nesting sea turtles that have already been affected by the oil.

In granting the emergency permit last week, the government recognized those worries. “There are a lot of doubts whether this is a valid oil spill-response technique,” Coast Guard Admiral Thad Allen, who heads the

GULF OIL DISASTER

No 'Smoking Gun' for Killer Oil

As turtle and dolphin corpses wash up on Gulf of Mexico beaches, scientists face a sleuthing challenge worthy of *CSI*: determining whether oil had a hand in the deaths.

In the past month, the stranding network operated by the National Oceanic and Atmospheric Administration (NOAA) has counted more than 200 dead sea turtles and more than 20 dead dolphins. Turtles normally strand during the summer nesting season, and the number of strandings in May has averaged 47 in the past 5 years. Although that's far below the current tally, none of the dead turtles had oil on its shell or skin, only one dolphin appeared to be oiled, and the 67 turtle necropsies that had been performed when *Science* went to

press had failed to find oil internally. However, officials are loath to either implicate or rule out oil as the culprit.

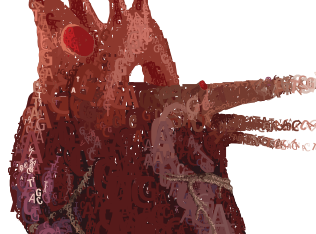
In addition to rescuing and cleaning oiled animals, NOAA is simultaneously gathering evidence for the government's Natural Resource Damage Assessment (NRDA), the estimate of a humanmade disaster's environmental impact that ultimately goes to court. Animals that died from exposure to oil factor into that estimate, says Michael Ziccardi, a University of California, Davis, wildlife veterinarian currently advising NOAA's efforts in Houma, Louisiana.

“The more environmental damage due to a spill, the higher the damage assessment is on a dollar basis,” Ziccardi says. “Animals do directly affect the amount.”

But pinning an animal's death to oil is tricky. Clues could be as obvious as oil on an animal's skin and in its mouth, eye irritation, or, during a necropsy, a tarball in the intesti-

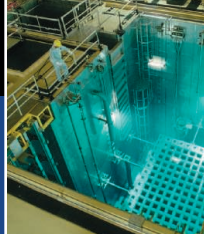


Victim. Just one of the dead dolphins found so far had oil on its body.



The case of the missing genetic risk factors

1220



Nuclear security technology

1222

national response, said at a press conference. “But we’re not averse to attempting this as a prototype.” The permit allows the state to build 72 kilometers of berm, less than half of its proposal. The federal government will front the bill for one section of the berm (with hopes of collecting from BP). The rest would have to be funded by the state, which could likewise try to get reimbursed by BP.

The scope of the permit may be expanded—and the federal government would consider funding more of it—if the project shows a “net environmental benefit,” according to Allen. It’s not clear what that means, but it could be as simple as checking whether the berms hold up and collect oil. “If these berms are keeping the beaches behind them clean, then that’s a pretty persuasive argument that they are doing something good,” Young says.

Still, experts doubt that the berms will catch much oil. For one thing, it could take many months to construct them, given that hurricane season has begun. And the 100-meter-wide berm will likely wash away quickly. Because the berm is steeper than the natural beach, waves erode it faster than normal. Even small storms could do serious damage and likely scatter the sand about the muddy sea floor where it couldn’t be retrieved. That would mean that much sand—a precious commodity needed for



coastal restoration—would be lost.

Coastal scientists are relieved that the project is starting out at half-scale and that the corps put the most plentiful and easily accessed sources of high-quality sand off-limits to dredging for now. The permit allows sand to be gathered only from Pass a Loutre, a wide channel of the Mississippi. “It’s probably quite good sand and could be replenished by the river,” says Denise Reed of the University of New Orleans. In addition, the dredging and bulldozing will have

to follow guidelines for avoiding harm to endangered species that live there, such as the piping plover.

Robert Twilley of LSU is hoping that the permit can be modified to allow sand to be placed not on a berm in front of the western barrier islands but rather on top, where it would be likely to escape erosion. “If we’re going to move sand, we should do that for the purpose of restoration, so we’re not building an artificial landscape,” Twilley says.

—ERIK STOKSTAD

nal tract and oil in unexpelled feces. But some clues show up only under a microscope.

Oil “causes profound effects in clinical pathology,” which vary based on the dosage and length of exposure, says Gregory Bossart, a veterinary pathologist and immunologist at the Georgia Aquarium in Atlanta who worked with oiled birds and turtles during the massive spills of the Persian Gulf War. The most toxic components of crude oil are polycyclic aromatic hydrocarbons (PAHs), volatile molecules that irritate every system in the body. Pathologists looking at mucous membrane tissue under a microscope might see inflammation, a sign of inhaling fumes. Ingested oil causes hemorrhaging in the gastrointestinal tract and damage to the liver and kidneys.

But even these microscopic clues aren’t the final word, Ziccardi says: “The same cel-

lular changes can be seen with diseases ... [or] starvation. That’s why a lot of this is very much like *CSI*, trying to put together the clues.”

On the other hand, “death tends to be a dynamic process,” Bossart says. Oil exposure can kill by making animals more vulnerable to other diseases, signs of which “could cloud the picture if you’re trying to see a simple cause and effect.” After the *Exxon Valdez* tanker spill, for instance, sea otters that had inhaled fumes developed pneumonia, he notes. “These are very complex morbidity and mortality issues that we need to be careful in how we interpret.”

Toxic PAHs are usually metabolized and flushed fairly quickly, and that makes them hard to use as a marker for oiling. But they can also accumulate in fatty tissue like ovaries and other reproductive tissues, says

NOAA environmental scientist and ecotoxicologist Simeon Hahn, who is leading the ongoing NRDA for sea turtles and marine mammals. This could be especially useful for determining oiling in animals too decayed to necropsy. His team will send tissue samples from all found animals to a laboratory to be analyzed for PAHs and their metabolites, he says. It could take weeks or months for those results to be announced.

Meanwhile, the unprecedented use of nearly 4 million liters of chemical dispersants has added a new wrinkle, forcing NOAA to develop protocols analogous to PAH testing to flag dispersant poisoning, something it’s never had to do before, and a scientific challenge, Hahn says: “In general, there’s not a good understanding of how the material accumulates in tissue.”

—LAUREN SCHENKMAN



Plane view. Mounted on a Boeing 747, SOFIA will fly at an altitude of 12,000 meters.

ASTRONOMY

Long-Delayed Airborne Observatory Takes First Flight, Sees First Light

It was 13 years in the making: a jetliner with a 2.5-meter telescope looking out into space through an opening on the rear left side of the aircraft. Four years ago, with NASA poised to yank funding, it looked as if this airborne observatory would be grounded forever. But last week, the Stratospheric Observatory for Infrared Astronomy (SOFIA) glided off the runway of the Palmdale Airport in California for its first observational flight, drawing cheers from scientists who expect it to open a new window on the nearby universe.

Designed to fly for 8 hours at a stretch

some 12,000 meters above Earth, SOFIA occupies a niche between ground-based telescopes, which cannot see infrared rays well because of atmospheric moisture, and space telescopes, which cannot be upgraded or serviced. Unlike Herschel, the European space observatory that has been snapping pictures at near-infrared wavelengths since its launch a year ago, SOFIA's Faint Object Infrared Camera (FORCAST) will operate in the mid-infrared, at wavelengths between 5 and 40 micrometers. That's a range in which the warm dust and gas of inter-

stellar clouds, planetary nebulae, and planet-forming disks around stars become visible.

FORCAST is just the beginning, says astrophysicist Eric Becklin, chief science adviser for SOFIA. In the next few years, eight more instruments currently in the pipeline will give SOFIA unique capabilities over a wider range of infrared wavelengths. As technology advances, SOFIA's potential will be limited only "by the imaginations of instrument builders," Becklin says, noting what SOFIA's managers describe as its key advantage over the competition.

SOFIA's first observational run on 25 May went smoothly, but the project itself has had a long and bumpy ride. Begun in 1996 as a collaboration between NASA and the German Space Agency, the project was initially estimated to cost \$250 million and to glimpse first light by 2005. But its cost quickly ballooned, in lockstep with missed deadlines, and a 2004 Government Accountability Office report cited SOFIA as an example of unrealistic baseline estimates typical of many astronomy missions. "We underestimated how technically difficult it is to put this very large hole in this very large aircraft," says Erick Young, the observatory's science missions director.

In 2006, NASA officials announced that the agency intended to pull the plug start-

CLEAN COAL

In Quest to Save Energy, China Ignores Simple Answer

BEIJING—At the Strategic and Economic Dialogue here last week, China and the United States found common cause on the need to team up on clean energy research. But one low-tech measure would allow China to make huge strides on its own in energy savings and pollution reduction: All it must do is wash its coal before burning it, says Xu Xuchang, a combustion expert at Tsinghua University here. But it doesn't, and this bodes ill for China's frantic efforts to meet a looming energy-efficiency target.

China's energy-conservation plan for 2006 to 2010 calls for a 20% reduction from 2005 levels in energy intensity—energy consumption per unit of GDP—and a 10% reduction in pollution. With time running out, China isn't close to achieving either target, and leaders are worried. In an all-hands-on-deck State Council meeting last month, Prime Minister Wen Jiabao urged local officials and executives of state enterprises to "eliminate excess production capacity with an iron fist," according to news reports.

That will not be easy. The last several years have been one step forward, two steps back for China's energy-savings effort. From 1980 to 2002, the country achieved a 5% average annual reduction in energy intensity. Then, fueled by explosive growth in automobile manufacturing and real estate, energy intensity rose 3.8% per year from 2002 to 2005. Over that period, the "Chinese economy drove a far greater energy demand than in any nation at any time," says Mark Levine, founder of the China Energy Group at Lawrence Berkeley National Laboratory in California.

As bad as things looked, China showed it could "turn the supertanker around," Levine says, by managing for a few years to reduce energy intensity. One bright spot is the "top-1000 energy-consuming enterprises" program, which allots energy savings quotas to companies. By the end of 2008, the program had achieved its 2010 target and saved 106 million tons of coal equivalent (MTCE) in energy. A second success story is the "large substituting small" scheme, in which 7467 outdated, noxious power-generation units were replaced with more efficient ones.

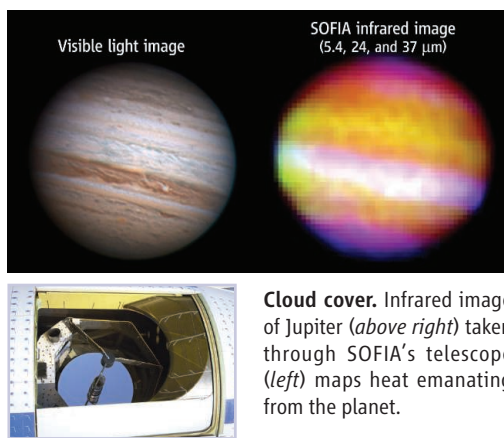


Out with the old. China has the muscle to demolish inefficient power plants but lacks the wherewithal to wash coal.

ing the following year. The decision sparked protests from the German collaborators and from SOFIA's supporters in the U.S. science community. "There was a large congressional effort to reverse the decision," says Becklin. NASA changed its mind and stuck with the project, even though the final bill came to exceed \$900 million.

The long delay in getting SOFIA up and running has frustrated some of the researchers involved in building instruments for the observatory. Reinhard Genzel, an astrophysicist at the Max Planck Institute for Extraterrestrial Physics in Garching, Germany, says his research group—which has built a far-infrared imaging spectrometer for SOFIA—will be ending its role in the observatory after handing over the instrument. "We were going to fly this instrument 4 years ago when SOFIA would have been the first observatory with this capability," Genzel says. "Now we are spoiled with Herschel data, so SOFIA is not as attractive." However, Genzel adds that SOFIA will have several other capabilities—such as high-resolution spectroscopy—that Herschel cannot match.

Scientists planning to use SOFIA are look-



Cloud cover. Infrared image of Jupiter (above right) taken through SOFIA's telescope (left) maps heat emanating from the planet.

ing ahead. The flight last week "was a success, showing that the vibrations of the telescope in flight are well within the predicted bounds," says Mark Morris, an astronomer at the University of California, Los Angeles, who has been allotted time on the observatory when it begins making science flights in October. Repeated launch delays quashed his earlier plans to observe the galactic center or the Orion nebula, Morris says, but now he is confident that he'll get his chance. "It may be a good time to study some nearby star-forming regions such as Taurus," he says.

—YUDHIJIT BHATTACHARJEE

However, rising energy demand has swamped those savings. China's \$586 billion stimulus package, says Levine, "particularly boosted construction and other energy-intensive activities," which "put a major crimp" in its conservation plan. Energy intensity shot up 3.2% in the first quarter of 2010 alone.

Hoping to squeeze more savings out of tried-and-true programs, the State Council has ordered enterprises to cut an additional 20 MTCE. But to achieve its long-term conservation goal, China must cap total energy consumption, argues Ni Weidou of the Tsinghua-BP Clean Energy Research and Education Centre. When policymakers set the target of 20% reduction of energy intensity, they assumed a total consumption of 3000 MTCE by 2020, says Ni. By 2009, consumption had reached 3100 MTCE.

Failure to rein in consumption imperils another goal: to increase nonfossil fuels to 15% of China's energy mix by 2020. Since 2005, the mix has hovered around 70% coal, 20% crude oil, 6% hydropower, 3% natural gas, and 1% nuclear, with negligible contributions from renewable sources such as biomass, solar, and wind. For the foreseeable future, coal will remain king.

Burning cleaner coal is critical, says Xu. Power plants and some 480,000 coal-fired boilers that generate steam for industry account for up to 85% of China's coal consumption. Because most of this coal is not washed or sifted, it is high in sulfur and ash and too powdery. The result is poor combustion that accounts for more than two-thirds of the country's sulfur dioxide emissions and the majority of particulate matter pollution. Burning cleaner coal would increase industrial boiler efficiency by roughly 15%, says Xu. However, concerned about their bottom lines, coal mines, the railroad monopoly, and the power industry all oppose regulating coal quality.

China must solve this kind of "governance problem" on its own, says Zha Daojiong, a political scientist at Peking University. In the meantime, he says, the nascent China-U.S. clean energy collaboration could focus on tasks such as how to monitor energy consumption at the factory level. Factories currently report their own data, which the central government has no way to verify. Any energy-efficiency claim therefore will be "just a numbers game," says Zha. And China is no closer to burning cleaner coal.

—HAO XIN

ScienceInsider

From the Science Policy Blog



The third time was a charm for members of the U.S. House of Representatives, which last week **passed a reauthorization of the America COMPETES Act** that would increase spending on research, science education, and innovation. <http://bit.ly/competes>

A top official at the National Institutes of Health told Congress that the U.S. government is moving quickly **to bring synthetic biology under review**. Anthony Fauci says two existing advisory panels will be broadened to encompass issues raised by synthesizing a genome. <http://bit.ly/syn-bio>

A report from the European University Association describes **six degrees of pain being felt by public institutions** in the wake of recent government spending cuts. <http://bit.ly/european-universities>

The European nations with the biggest combined responsibility for an ambitious fusion reactor project are still **scrambling to find a way to absorb its ballooning cost** in advance of an important meeting of ITER's governing council. <http://bit.ly/iter-budget>

A mathematician and a neuroscientist are millionaires thanks to **this year's Shaw Prize**, which divided its third \$1 million award among three astronomers. <http://bit.ly/shaw-prizes>

President Barack Obama's plan to reorient the U.S. space program took another drubbing on Capitol Hill last week, as lawmakers in the House of Representatives endorsed the sharp criticism from Apollo astronauts about the **Administration's plan to gut Constellation** and questioned whether NASA would ever receive enough money to fund the proposed new direction. <http://bit.ly/obama-nasa>

Rules issued last year by the National Institutes of Health don't eliminate **the need for guidance on stem-cell research** that is off-limits to U.S. government-funded labs, says a National Academies' panel in its final report on human embryonic stem cells. <http://bit.ly/last-hurrah>

See the full postings and more at news.sciencemag.org/scienceinsider.

ETHICS

Researchers to Return Blood Samples to the Yanomamö

The closest Kenneth Weiss has been to the Amazon rainforest was when he passed over it in an airplane last year. But the geneticist at Pennsylvania State University, University Park, says his freezers hold hundreds of 40-year-old blood samples from the Amazon's Yanomamö Indians. Now, in an agreement being worked out by Brazil, he and others are pulling tissue samples out of storage and preparing to have them shipped back to the jungle.

Weiss says he accepted the vials years ago as a favor to his postdoctoral adviser James Neel, who was retiring and wanted them preserved. Along with cultural anthropologist Napoleon Chagnon, Neel collected the samples from the Yanomamö in Brazil and Venezuela during fieldwork in the 1960s and early 1970s, and they've been stored since then in labs around the United States. (Neel died in 2000.) Weiss and others will be releasing parts of their collections to the Brazilian Embassy in Washington, D.C., which in turn will escort them back to Brazil and the Yanomamö tribe. Venezuela has not asked for samples taken from its Yanomamö tribes, Weiss says.

The return marks at least the third time that an indigenous group has retrieved DNA or other tissue from scientists, suggesting a shifting landscape in genetics studies on indigenous people.

Unlike a recent case involving the Havasupai tribe in Arizona's Grand Canyon (*Science*, 30 April, p. 558), those who worked with the Yanomamö have never fought the tribe's requests. The Yanomamö began seeking the samples about 10 years ago after the publication of *Darkness in El Dorado: How Scientists and Journalists Devastated the Amazon*. The controversial book by investigative journalist Patrick Tierney charged that Chagnon and Neel acted unethically in their dealings with the Yanomamö (*Science*, 19 January 2001, p. 416). The logistics of shipping the samples back has taken years to sort out.

"We can't just put a Paddington Bear tag on there and say 'Please deliver to Amazonia,'" says Weiss. Furthermore, there's confusion among Brazilian officials and researchers over what the Yanomamö will do with the material—in particular, some have said,



Back to the jungle. Blood samples collected by the late James Neel, shown here drawing a sample in 1968, and his colleagues will be returned to tribe members.

whether they might ingest the samples in a ritual. Chagnon considers that unlikely; the Indians "do consume the ashes of deceased kinsmen after cremating them," he says, but that's just ground-up bone. Still, the concern that the Yanomamö "could get sick from a new disease" is real, Chagnon agrees.

Researchers and diplomats alike want to ensure that the samples are safe and free of contaminants. That's easier said than done. The usual approach—heating material at very high temperatures—would cause the vials to explode. A suggestion to sterilize some samples with bleach was rejected, says Karen Pitt, special assistant for biological resources at the National Cancer Institute (NCI), which holds 477 vials. NCI is investigating the possibility of irradiating them. "We'd like to accelerate this," says Pitt.

Weiss agrees: "I'd rather have the freezer space and have these samples off my back." He received several hundred Yanomamö vials a few weeks ago from a colleague at Binghamton University in New York state in efforts to combine them for return. Both

Weiss and NCI ruled out further research on the samples when they learned that the Yanomamö wanted them back.

Scientists are increasingly trying to accommodate demands from indigenous groups. Three years ago, the Canadian Institutes of Health Research in Ottawa released new recommendations for aboriginal research requesting, among other things, that research be of benefit to the community, that researchers translate their publications into the language of the community, and that researchers get consent before transferring samples to a colleague.

"If you have a sample in your lab, you have been loaned it, you haven't been given it," says Laura Arbour, a medical geneticist at the University of British Columbia, Vancouver, in Canada who helped craft the Canadian guidelines. Arbour, who works with Canadian aboriginal populations, believes they should be treated as collaborators and shown drafts of papers prior to publication, something she routinely does in her own genetics work.

"I don't object" to this approach in principle, says Kenneth Kidd, a population geneticist at Yale University, but it would make research "a lot more difficult." He and his wife, Judith Kidd, have amassed 3000 samples from 57 populations over the years. It would be virtually impossible to find a nomadic tribe from whom samples were collected a decade ago and share a planned publication, he says.

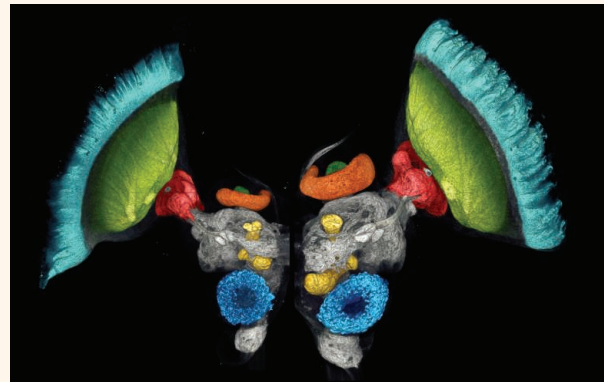
Regarding the Yanomamö, there was no question that the request to return samples would be honored, says Pitt. "When the specimens were originally collected, there wasn't anything called informed consent," she says. "Now we're becoming a lot more sensitive to the rights of study participants to withdraw from a study."

As for Chagnon, he vigorously defends his work and believes that the Yanomamö are being pressured to make demands by politically motivated nongovernmental organizations. Ultimately, he says, the people of the Amazon will lose out by retrieving and presumably destroying the vials collected so long ago, which might have been useful in biomedical research. "The Yanomamö, like all populations, depend on medicines," he says. "I think it's reasonable that they should participate in what it takes to develop medicines, instead of being just the recipients."

—JENNIFER COUZIN-FRANKEL

From *Science's* Online Daily News Site

Swarming Locusts Grow Big Brains. When desert locusts (*Schistocerca gregaria*) become too great in number, they transform from solitary insects into swarming crop raiders. Their brains also get a lot bigger, as revealed by computer-assisted laser microscopy. Not only are the brains of gregarious locusts (right) 30% larger than those of solitary locusts (left), gregarious locusts have proportionately larger regions that deal with higher brain functions (yellow and orange), such as learning and memory, researchers report in the *Proceedings of the Royal Society B*. This may help them survive in densely packed groups, where competition for food is so fierce that cannibalism occurs. <http://bit.ly/locusts>



Life Insurance for Ebola Scientists

When a German scientist accidentally pricked herself with a needle containing the Ebola-Zaire virus last year, scientists around the world tried to determine what the best course of action would be. Eventually, she agreed to take an experimental Ebola vaccine that had been shown in animals to offer about 50% protection, even when administered after exposure to the virus. And she lived.

But now, scientists led by virologist Thomas Geisbert of Boston University School of Medicine say they have something better for such mishaps: a new therapy based on small strands of RNA that interfere with the production of viral proteins. In lab studies, it offered complete protection to macaque monkeys when given daily for almost a week after an Ebola injection that would otherwise be fatal. The new treatment, reported in *The Lancet*, could also be used during Ebola virus outbreaks, the researchers say. <http://bit.ly/Ebola-therapy>

OCD? Your Immune System Could Be to Blame

Some people just can't help themselves. They wash their hands over and over, lock and relock doors, or obsessively rearrange the contents of their closet. Now, a study of mice suggests that the immune system, rather than the nervous system, prompts such compulsive behaviors.

Mice with a defective version of the gene *Hoxb8* overgroom themselves, tear out patches of fur, and give themselves sores—a condition akin to trichotillomania, in which humans compulsively pull out their own hair.

But when researchers led by molecular geneticist Mario Capecchi of the University of Utah School of Medicine in Salt Lake City went looking for *Hoxb8*-expressing cells in the brain, they didn't find neurons; they found

microglia, immune cells that are often born in the bone marrow but that can migrate to the brain. What's more, transferring marrow from *Hoxb8*-lacking mice into healthy rodents provoked compulsive grooming, the researchers report in *Cell*. And when mice deficient in *Hoxb8* received marrow from healthy animals, they cut back on their ablutions.

The results, say the researchers, raise the possibility of treating obsessive-compulsive disorder by targeting the immune system rather than the brain, perhaps by developing drugs to alter microglia activity. <http://bit.ly/OCD-mice>

Does a Country's Dirt Determine Its Destiny?

Chad is dirt poor because its dirt is poor. Germany is relatively rich because its soil is rich. That's the provocative conclusion of a new study, which suggests that just two fundamental factors—soil type and climate—can largely explain why humans have prospered in some places but not in others.

Jan Beck, an insect ecologist at the University of Basel in Switzerland, and his student

Andrea Sieber selected 19 climate and soil variables that could influence how humans use land, including soil type and precipitation in the coldest and warmest seasons. Then, using computer models originally developed to predict the ecological niches best suited to certain insects, they calculated how well those variables predicted common modes of human land use in Asia, Africa, Europe, and Australia.

The resulting maps, which appear in *PLoS ONE*, show that areas fertile enough for crop farming are able to support much denser and richer populations than regions more suited to ranching or hunting. Much of Europe, for example, turned out to have a relatively farming-friendly climate and landscape, perhaps explaining why it has prospered. But, Beck notes, his model failed to correctly predict present land uses up to one-third of the time, suggesting that history and culture do influence how humans use the land. <http://bit.ly/Dirt-is-Destiny>

Read the full postings, comments, and more at news.sciencemag.org/sciencenow.



Violent Galaxy Collisions Power Mega Black Holes

Among the most luminous things in the universe are active galactic nuclei (AGN): gigantic black holes that can emit as much energy as 10 billion suns. Why these objects are so much brighter than ordinary supermassive black holes has long puzzled researchers. Now, with the aid of NASA's orbiting Swift x-ray observatory, astronomers have confirmed what theorists consider the likely explanation: AGN derive their stupendous power from the merger of two galaxies. "We think we have the 'smoking gun' for merger-triggered AGN," says graduate student Michael Koss of the University of Maryland, College Park, who, with colleagues, reports the finding in *The Astrophysical Journal*. <http://bit.ly/mega-blackholes>



Major Heart Disease Genes Prove Elusive

So far, genome-wide association studies have not found common genes with a big impact on heart health; researchers hope that the low-effect genes they are finding will help identify pathways and drug targets

THE EXCITEMENT BEGAN 5 YEARS AGO, when a study of 146 Caucasian volunteers turned up a common gene variant among those with the eye disease macular degeneration. Researchers had used a new strategy: They scanned large stretches of the genomes of the sick and the healthy and found a single DNA base that was much more likely to be present in those whose eyes were failing.

The finding was remarkable: Relatively few people participated in the study, yet those

with two copies of the suspect gene variant had 10 times the risk of macular degeneration, a huge increase. Furthermore, the method the group used, called genome-wide association (GWA), had some big advantages: It was unbiased, testing thousands of gene-disease associations at once, not just a researcher's favorites. And it pointed to common variants, found in at least 5% of individuals studied. GWA studies offered hope of identifying people at risk for diseases, uncov-

ering new disease mechanisms, and finding new targets for therapy.

Almost immediately, researchers applied GWA to other conditions. But they were quickly stymied. "People did studies with 300 or 500 people and didn't find anything, then did 1000 and didn't find anything," says Deepak Srivastava, who directs the Gladstone Institute of Cardiovascular Disease at the University of California (UC), San Francisco. It quickly became clear that macular degeneration was an exception. Most GWA studies needed 10,000 or more volunteers to get a statistically significant result, because the effect of each gene was so small.

Since the human genome was sequenced 10 years ago, technology has moved with lightning speed; many now believe that GWA methods, which cover a fraction of the genome, are becoming obsolete. Sequencing costs continue to plunge, and within a few years sequencing entire genomes of hundreds of subjects will be financially feasible.

What has the GWA experience taught us? The results from one group of GWA studies, for heart disease, are typical, with a mixed record and an uncertain legacy. The technique has identified dozens of variants, but all have weak effects; so far, almost none has led to DNA changes that actually cause disease. Researchers have had more success finding variants that link to tightly defined conditions like high cholesterol than to heart failure, a catch-all disease.

"At the end of the day, we have a bunch of loci and genes, but none of them" do all that much to raise the risk of heart disease, says Eric Topol, a cardiologist and director of the Scripps Translational Science Institute in San Diego, California. Nor have they yet altered our understanding of how the heart falters—knowledge, Topol says, that will take time to develop.

GWA studies still have many backers. "We have new technology that's enabled us to look at things we've never seen before," says Bruce Psaty, a cardiovascular disease epidemiologist at the University of Washington (UW) School of Medicine in Seattle. And Francis Collins, director of the National Institutes of Health (NIH), has said that the approach has provided "1000 new drug targets" (*Science*, 28 May, p. 1090).

Clues missing

The first GWA results for heart disease hit in 2007. Three studies examined coronary artery disease, in which plaque builds up in the arteries and narrows them. Together with subsequent studies, they identified 12 new genetic variants, called single-

nucleotide polymorphisms (SNPs). Intriguingly, only about four had an apparent connection to known heart disease risks, like high cholesterol. “We do not have any clue why these SNPs are related to coronary artery disease,” says Jeanette Erdmann, a human geneticist at the University of Lübeck in Germany, who participated in one of the studies.

One advantage to the GWA studies was that unlike much earlier work in common diseases, scientists could easily repeat the findings, suggesting that the results were real. But there were disappointments, too. In the macular degeneration study, two copies of the culprit SNP raised the risk by 1000%. But in the GWA heart studies, carrying two variants increased the risk of disease in older adults by, at most, 60%. This makes logical sense: Variants that dramatically boost disease risk aren’t likely to be common because they would have been removed from the gene pool by evolution.

But the results seem predictable partly because we’re viewing them in retrospect. “We had no clue” before beginning this work how potent common disease DNA would be, says Psaty. “Our expectations about the effect sizes were more hopeful than they should have been.” To date, GWA findings cannot be used to predict risk any better than family history for most chronic diseases. Furthermore, they are only a very small piece of disease heritability: In coronary artery disease, the 12 variants found represent “well under 10%” of the total heritability of that condition, says Nilesh Samani, a cardiologist at the University of Leicester in the United Kingdom, who helped lead one of the studies.

Getting at biology

While GWA data aren’t very useful for calculating risk, GWA backers are hanging on to another hope: that they can elucidate disease biology. The big hurdle here is lack of precision. A new SNP “doesn’t give you cause, it gives you zip code,” says Gerald Dorn, a cardiologist and director of the Center for Pharmacogenomics at the Washington University School of Medicine in St. Louis. That’s because the SNPs identified in these studies aren’t themselves driving disease—rather, they travel alongside genes or other variants that do.

To find those driver genes, biologists must pounce on GWA results and use them to track down disease-causing variants. “The molecular biologists are not wild” about doing this, says André Uitterlinden, a molecular geneticist at Erasmus University Medical Center in the Netherlands. It isn’t

easy. Sometimes the SNP is “in a desert,” Uitterlinden says, with no genes nearby; other times, “you hit upon a very gene-rich area” and it’s hard to know where to turn.

Undaunted, some groups are pursuing GWA variants. In coronary artery disease, a region called 9p21 is getting attention. It’s among those most strongly associated with heart disease risk and has been linked to abdominal and brain aneurysms as well. In March, a group of researchers reported in *Nature* that when they knocked out 9p21 in mice, arterial smooth muscle cells proliferated, something also seen in atherosclerosis. The DNA behind the effect still hasn’t been pinned down, says Ruth McPherson of the University of Ottawa Heart Institute, one of those who pinpointed 9p21, but she’s hopeful the variant will offer “new biological insight” into artery disease.

While GWA studies in atherosclerosis or heart failure have yielded relatively few “hits,” and none that strongly increase risk, geneticists have had more luck with narrowly defined physical characteristics. “We had 23 hits in the red cell traits,” such as red blood cell count, and “only one for stroke,” says Psaty. Cholesterol levels have garnered about 80 hits. These are “closer to the biology of the genome,” says Uitterlinden. That’s very different from a condition like heart failure or stroke, which he describes as “the waste bin of all your bad” traits. There are many roads to heart failure, but not so many to red cell numbers.

Psaty and Uitterlinden are part of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, a group in North America and Europe. It combines several massive cohort studies, including the Framingham Heart Study and the Rotterdam Study, to perform GWA studies in heart disease, aging, and related conditions. Various teams, including CHARGE, have been conducting meta-analyses on existing GWA studies; one paper soon to be submitted, researchers say, will roughly double the number of variants behind coronary artery disease. Each of the new gene regions identified by the variants will be even less of a factor in heart disease than existing ones.

But researchers hope the low-effect genes will coalesce on new molecular pathways—and, eventually, enable them to accomplish what GWA studies haven’t done so far: refine an individual’s risk of disease. “It might be possible to have a SNP chip with perhaps 100 common variants that will give information” about risk of heart disease, says McPherson. But in the

new, cautious language of GWA work, she adds that she doesn’t “want to be too optimistic about that—but it’s possible.”

Optimism is more apparent around drug targets. Many point out that a gene variant’s effect on risk doesn’t necessarily correlate with what modifying it with a treatment can do for disease. “The main thing that was delivered [by GWA studies] was a treasure trove of candidates,” says Charles Langley, a population geneticist at UC Davis. “The distance of that path is somewhat depressing to people who are at the front line,” such as those who treat patients, “but it’s a huge boon to basic researchers.”

Alternative approaches

Dissatisfaction with GWA work continues to fester, however. Skirmishes have broken out among geneticists. In April, prominent geneticist Mary-Claire King of the UW School of Medicine wrote an essay in *Cell* lambasting GWA studies. She and her UW colleague Jon McClellan posited that GWA studies have failed to explain most heritability of common diseases. King’s paper was swiftly attacked by geneticists on various blogs, who defended GWA work and argued that it has revolutionized our understanding of disease (<http://bit.ly/GWAS-Heart>).

Underlying King’s and other critiques, some say, is whether precious time and money might be better spent searching for rare variants with potentially stronger effects. “GWA [work] has ruled genetics for 5 years,” and those researchers “have gotten all the money,” says Dorn, who’s done some himself.

The pendulum is now swinging in a different direction. NIH is pouring money into exome sequencing, in which all parts of the genome that code for proteins are sequenced—allowing biologists to dig even deeper into genetic variation.

It is a bit illogical to quarrel over rare variants versus common ones, say many researchers, because both are important. The fight may fade in the coming years, as sequencing full genomes becomes financially feasible—a strategy everyone will turn to, says Dorn. And GWA studies might help in years to come thanks to the vast population databases they’ve built up, Srivastava notes.

Still, we’re a long way from where we want to be, he believes. “Can we find loci that would teach us about the mechanisms of disease?” Not yet. “In retrospect,” Srivastava believes, GWA “wasn’t worth the expenditure—but that was hard to say prospectively.”

—JENNIFER COUZIN-FRANKEL

NONPROLIFERATION

An Unending Mission to Contain The Stuff of Nuclear Nightmares

With new tools and painstaking analyses, researchers are helping to detect clandestine nuclear-weapons research and keep deadly material from going astray

LOS ALAMOS, NEW MEXICO—Just after midnight on 8 November 2007, four armed men broke into the Pelindaba Nuclear Research Center near Pretoria in South Africa, disabling a 10,000-volt electric fence guarding the perimeter. Breaching several layers of security, the intruders made their way to the emergency control room, shot an employee, and stole a computer. Closed-circuit cameras caught their images, but security personnel failed to notice. While the attack was going on, another group made an unsuccessful attempt to break into the center.

The incident sent a chill down the spines of counterterrorism officials around the world. Even though the attackers did not appear to have planned to steal any of the weapons-grade uranium stockpile stored at the facility, the fact that they were able to spend 40 minutes inside showed how easily terrorists might strike nuclear facilities like Pelindaba.

Fears of terrorism have added new urgency to the goal of preventing the proliferation of nuclear weapons, which was the subject of international deliberations held in New York City last month to review the Nuclear Nonproliferation Treaty. The task at the core of nonproliferation efforts, however, remains what it has been for decades:

keeping track of nuclear materials and sites to ensure that countries do not make bombs, either in secret or under the pretext of developing nuclear power. It's a daunting scientific and technological challenge.

The International Atomic Energy Agency (IAEA)—the world's nuclear watchdog—has traditionally met that challenge by dispatching inspectors to nuclear sites to verify the designs of facilities and carry out measurements of materials at different points of the fuel cycle. But nonproliferation experts acknowledge that this system of spot checks is far from foolproof. It leaves open the possibility of covert activities between inspections, such as enriching uranium to a grade high enough for bombmaking or diverting material in quantities small enough to evade inspectors' measurements.

That's why in recent years, IAEA has sought new safeguards that would enable the agency to monitor nuclear facilities remotely around the clock, keeping tabs on equipment and materials with greater assurance than before. These technologies include advanced laser-based surveillance systems and a variety of detectors to track nuclear materials continuously—with

unprecedented precision and certainty—as they move through a facility. So far, only a handful of sites have deployed the new systems, but officials expect them to be put into wider use in the coming years as more nuclear reactors are built around the world, straining IAEA's inspections capacity. "The IAEA is very concerned about the manpower challenge it will have to face if even a fraction of the anticipated nuclear renaissance becomes a reality," says Philip Hypes, who coordinates research projects for IAEA

at Los Alamos National Laboratory (LANL) in New Mexico.

Monitoring facilities may be a central piece of the nonproliferation puzzle, but it's not the whole story. Over the past decade,

IAEA has begun to search satellite images and the scientific literature for clues about covert nuclear activities. The work goes far beyond material accountancy, says Brian Boyer, a nuclear engineer who once worked as an IAEA inspector in Europe and now leads the nonproliferation team at Los Alamos. "We also want to know what countries might be doing in their research labs and elsewhere, if they might be building up a process or doing an experiment in a clandestine facility."

Online

sciencemag.org



Podcast interview
with author

Yudhijit Battacharjee.

Nuclear blues. Glow from fuel rods in a storage pond tells of plutonium, but not how much.

Added to proliferation concerns are fears about terrorists acquiring nuclear materials to fashion a conventional bomb or a less-potent radiological device. To address that threat, in April the U.S. government and 46 other countries announced the intent to secure all nuclear materials in their possession within the next 4 years. Experts say that's a tall order given the weak security measures at many nuclear facilities; improved safeguards will no doubt help but are only part of the solution. "A dirty little secret is that safeguards have very little to do with safety or with guarding," says Matthew Bunn, a nuclear policy researcher at Harvard University. "Safeguards themselves are not going to stop a gang of thieves from making off with nuclear material." Old-fashioned security is vital, he says.

Weak links

On display at the Bradbury Science Museum, down the road from LANL, are replicas of Little Boy and Fat Man—the atomic bombs that destroyed Hiroshima and Nagasaki. Both were designed at Los Alamos, whose primary goal continues to be maintaining a reliable U.S. nuclear deterrent. But since the 1960s, the lab has also spearheaded the development of safeguards aimed at preventing another Hiroshima. "It is an enduring mission that has become increasingly important," says William Rees, principal associate director for global security at LANL. "With commercial reactors making a comeback, it's vital that we have foolproof systems in place to insure that the nuclear materials don't fall into the wrong hands."

Some of the latest monitoring technologies developed at the lab are being tried out at a uranium-enrichment plant and a reprocessing plant in Rokkasho, Japan. In May, LANL researchers delivered a new system for measuring the enrichment level and quantity of uranium in 2-meter-long steel cylinders moving through the Rokkasho enrichment facility (see photo, p. 1224).

Currently, inspectors estimate the proportion of fissile uranium-235 in the material (usually uranium hexa-



"With commercial reactors making a comeback, it's vital that we have foolproof systems in place to insure that the nuclear materials don't fall into the wrong hands."

**—WILLIAM REES,
LANL**



Out of circulation. Los Alamos team picks up used radioactive material from a U.S. hospital, as part of the lab's worldwide program to track down and recover "orphan" sources.

fluoride) inside a cylinder by using portable detectors to measure gamma rays coming from the container's outer layer of content. To determine the overall amount, they simply weigh the container on a scale. If a facility wanted to divert enriched uranium out of such containers into a secret weapons program, it could fool inspectors by replacing some of the uranium in the container with another material such as lead and lining the inner wall of the cylinder with uranium to emit gamma radiation. The container could then, in theory, give the appearance of being filled entirely with uranium.

The new system—which consists of tubes wrapped in cadmium and filled with helium-3—counts neutrons that fluorine atoms in uranium hexafluoride emit when bombarded by alpha particles from the natural radioactive decay of uranium. Because neutrons—unlike the weak gamma rays uranium emit—can penetrate dense material, they can make it out from within the core of the container, providing a measurement of the entire content.

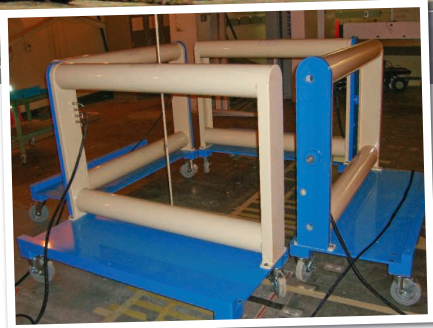
Developed by Martyn Swinhoe, Karen Miller, and others at LANL and other

institutions, the system is being installed in the corridor between the facility's storage area and the enrichment plant's entrance to enable the remote measurement of each cylinder entering or leaving the plant. In the past few years, Swinhoe and his colleagues have helped outfit Rokkasho with similar systems to keep track of plutonium in canisters and nuclear waste.

Scientific techniques, however, still can't keep tabs on nuclear material throughout the entire fuel cycle. The cycle starts with the conversion of mined uranium ore into uranium hexafluoride and ends with the reprocessing of spent fuel to retrieve plutonium, one of the products of burning uranium and a bombmaking ingredient itself. One major challenge is quantifying the amount of plutonium in spent fuel assemblies—arrays of metal rods that are typically stored in cooling ponds, where they continue to generate heat for years.

Currently, inspectors take stock of spent fuel by peering into the cooling pond through special goggles to look at Cherenkov radiation—a blue light the water gives off as it is excited by electrons spewing from the rods. The method is a crude way of confirming that the assemblies in the pond are indeed spent fuel rods. "You could fool inspectors by swapping out a few rods in an assembly with rods of depleted uranium and divert them to a covert reprocessing plant," says Boyer.

That's why researchers are seeking more direct approaches to measurement. In one



Keeping tabs. Japan's Rokkasho reprocessing plant is testing new equipment (inset) for measuring the amount of uranium it handles.

technique, they count neutrons that the rods emit both spontaneously and through induced fission. The spontaneous emissions come mainly from curium—one of several radioactive elements that reside in the spent rods along with plutonium. Researchers first measure the neutrons the rods emit naturally. Then they bombard the rods with external neutrons to induce fission in the plutonium, which causes an additional emission of neutrons that corresponds to the amount of plutonium present.

Other groups have been developing new surveillance technologies for nuclear sites. At the European Commission's Joint Research Centre in Ispra, Italy, Vitor Sequeira and his colleagues have designed a system that uses lasers to scan entire rooms within the plant to create detailed three-dimensional reference maps of the equipment. By comparing scans taken at different times, inspectors can check whether the equipment design has been modified even slightly. "Unlike traditional surveillance cameras," Sequeira says, "the system can detect millimeter changes" such as those that would result if the piping had been subtly altered to divert small amounts of nuclear material.

Words to the wise

Researchers at LANL and elsewhere are helping to advance nonproliferation not just through technology development but also by sleuthing. Increasingly, they are scrutinizing scientific and technical papers for evidence of undeclared nuclear

research and development that could help with secret weapons-building plans. For example, says Richard Wallace, a former IAEA analyst who now works at LANL, "you could have a country that says it is not working to reprocess spent fuel from nuclear reactors—and then you may go into the literature and find papers discussing efficiency of the methodology for separating isotopes from spent fuel."

That kind of digging has yielded results in the past. In 2004, for instance, IAEA officials combing through the proceedings of a 1995 conference on nuclear fuel-cycle systems in Versailles, France, spotted a paper by three Egyptian researchers. The publication—*Kinetic Separation of Uranium from Selected Fission Products*—indicated that the Egyptian Atomic Energy Authority in Cairo might once have done work on reprocessing that it had not declared to IAEA. Similar literature-mining led researchers to discover a dozen papers between 1991 and 2004 showing that scientists at the Korea Atomic Energy Research Institute in Daejeon had conducted uranium-enrichment experiments without declaring them. In both cases, IAEA ultimately found no evidence of evil-doing, but both countries were forced to explain why they had not reported the work.

The absence of publications in a declared area of research can also be a warning sign. "If there has historically been activity on a research topic, and you see that the publication record of the country's universities in that area has been high or steady—and then it suddenly drops to nothing—you wonder about the change," says Wallace, refusing to cite specific examples. "What could those people be doing now?"

Scavengers and thieves

Deception isn't the only way material can go astray. As the Pelindaba attack showed, even good safeguards can still leave nuclear sites vulnerable to pillaging. "Many research reactors actually have lower levels of security than Pelindaba," says Harvard's Bunn. To tackle the threat of nuclear terrorism, he says, governments need to invest more on fortifying nuclear sites—in part by developing tighter security systems.

Researchers are working on them. At the Joint Research Centre in Ispra, for example, Marco Sironi and colleagues have developed special seals for spent fuel assemblies submerged under water. Each seal consists of a set of overlapping metal discs with holes and cuts that give it a unique sonic signature detectable by a device that shoots ultrasound pulses at the seal and analyzes the reflected signal. In April, several of the seals were put in place to secure spent fuel at the Karachi Nuclear Power Plant in Pakistan—a country where growing Islamic extremism poses a high risk of nuclear terrorism.

Terrorists could also strike with a dirty bomb made using radioactive materials collected from the garbage cans of hospitals. "Dirty bombs may not cause the same loss of life as a nuclear device, but they would cause severe disruption, and if detonated in a metropolitan area, they would cause enormous economic damage," says LANL's Julia Whitworth, who oversees a program for the recovery of orphan radioactive sources. The program, started a decade ago, is an important part of the broader initiative to achieve nuclear security, she says. Teams of scientists and technicians travel around the United States and overseas to pick up discarded radioactive sources—surgical gamma knives from hospitals, oil-well loggers buried in backyards, and the like. Earlier this month, Whitworth's colleagues flew to Manhattan to go through all the sources at a hospital that was shutting down for good. Whitworth says the United States has recently stepped up efforts to round up radioactive material in old x-ray equipment that other countries imported from it. "We just got a big shipment from Chile," she says.

As the international community strives to achieve nuclear security in 4 years, scientists will play an increasingly important role in the task, says David Albright, a physicist who runs the Institute for Science and International Security in Washington, D.C. In a world moving toward greater use of nuclear power, he says, the timely detection of proliferation activities "may well hinge on the clever use of scientific and technological innovations."

—YUDHIJIT BHATTACHARJEE

CREDITS (TOP TO BOTTOM): AFP/GETTY IMAGES; (INSET) LOS ALAMOS NATIONAL LABORATORY

Downloaded from www.sciencemag.org on June 3, 2010



UKRAINE

Renewing the Post-Soviet Steppe

Ukraine's steppe has largely disappeared, but researchers, conservationists, and farmers are devising plans to restore the vital grasslands while making money

Ukrainian ecologist Oleksander Mykytiuk launches Google Earth on his computer and zooms in on a satellite image of the eastern corner of his country. He points to a patchwork of highlighted areas in the rolling farmland of the Luhansk province. These were once fields of wheat or barley; now weeds fill the abandoned soil. The image shows the collapse of Soviet-era collective farming, but also, he hopes, Ukraine's new ecological opportunity. It could be a chance to restore some of the vast grasslands, or steppe, that Soviet agricultural planners destroyed. These grasslands, now scarce in Ukraine, are home to hundreds of threatened species, including grasses, insects, ground-nesting birds, and small mammals such as marmots.

Mykytiuk recently helped lead a \$3 million effort, funded by the European Union, to find ways of regrowing the steppe while capturing economic benefits from it. That research has now spawned a small but ambitious commercial venture—a Dutch-Ukrainian company that plans to fatten beef cattle on thousands of hectares of restored grasslands. Yet bringing back Ukraine's steppe faces obstacles, including competing conservation interests and widespread public disinterest. "People don't understand that the steppe is important for our country," says Tatiana Sova, director of steppe reserves in the Luhansk region.

Grasslands once dominated the eastern half of Ukraine, Europe's second-largest country after Russia. Most of them were plowed up in favor of crops, especially during the Soviet Union's campaign to expand food production from collective farms. Only small fragments survive, mainly in areas too steep or rocky to plow—or in about 40,000 hectares of protected nature reserves, such as the ones Sova oversees.

Ten years ago, Ukraine abolished state-run farms and divided up the land among their

workers. The government ignored advice that it set aside 8 million to 10 million hectares for restoration of steppes and forests, notes Vasily Kostyushin of the Ukrainian Academy of Sciences' Institute of Zoology.

Big agribusiness companies quickly snapped up the most productive land, leasing it from groups of owners. Large swaths of less valuable land, however, amounting to perhaps 10 million hectares in Ukraine, simply lie fallow.

This is the land that Mykytiuk and his colleagues are targeting. But bringing it back will demand more than just reseeding native flora or creating new reserves. "The problem is, the steppe isn't stable unless it is also consumed" by grass-eaters such as bison, marmots, and insects, says Vasily Tkachenko of the Institute of Botany in Ukraine's Academy of Sciences.

Tkachenko cites as evidence the deterioration of Ukraine's officially protected steppe reserves, where grazing is strictly limited and fires have long been controlled. As a result, a layer of decomposing grass built up on the ground, suppressing some steppe plants. Shrubs and small trees proliferated. Marmots and some ground-nesting birds started to disappear. Restoring Ukraine's steppe will demand more active management, contends Mykytiuk. Even then, people will only do it on a large scale if it's profitable, he says: "We need a business approach; an economic approach."

This is where the new Dutch-Ukrainian venture comes in. Ivan Kapatsiy, a farmer near the town of Bilovodsk, northwest of Luhansk, and Rieks Bosch, a Dutch environmental consultant, plan to raise high-quality beef

Rare breed. Red Steppe cattle (*left*) may help restore grasslands and prevent shrubs from taking over, as on a steppe reserve near Luhansk (*right*).

cattle on 5000 hectares of abandoned land near Luhansk. Kapatsiy has already leased the land. Some is former cropland that suffered from overuse, but a large section was once sheep pasture and was never plowed.

The next step is acquiring the money-making animals, which has proved complicated. Cattle breeds that offer top-notch beef and can thrive in Ukraine's continental climate are rare, Bosch says. Only about 400 individuals remain of one such breed, known as Red Steppe or Red Ukrainian. They survive on Ukraine's largest steppe reserve, Askania-Nova.

Kapatsiy wants to start with at least 100 animals, but so far the venture's lenders have provided enough money to acquire only 25. Bosch is now reworking the business plan. Down the road, he says, carbon-trading could sweeten the pot; restoring the steppe with extensive grazing would capture carbon in the soil, so farmers could sell carbon credits on European carbon markets. The Dutch Ministry of Agriculture, Nature and Food Quality has approved \$150,000 in research funding for a study of the farm's economic viability and its impact on the soil and wildlife.

Bosch says that local officials in Luhansk are supporting the effort, because a revitalized steppe, with marmots standing sentry beside their holes, could become a center of recreation and local tourism. Such official enthusiasm for steppe restoration is rare. Most people, Mykytiuk says, think of the steppe as a kind of void, without the value of cropland or even forests.

Ukraine has a State Committee on Forestry devoted to managing and promoting forests, but nothing similar exists for the steppe. The forestry agency, in fact, has competing plans, proposing to plant trees on 465,000 hectares of the country's degraded or abandoned lands. According to the National Ecological Centre of Ukraine, an environmental advocacy group, foresters have already ripped up remnants of native steppe to plant saplings. Such plans make Sova furious. "The steppe is always in danger, and this forest program could be the last straw," she says.

—DANIEL CHARLES

Daniel Charles is a writer based in Kyiv.



Eying profits. Farmer Ivan Kapatsiy.



HYDROLOGY

Along the Indus River, Saber Rattling Over Water Security

Pakistan accuses India of contravening a treaty that governs Indus water; with climate change likely to make things tenser, better data sharing may forge mutual trust

ISLAMABAD—High in the Himalayas, near the militarized zone that divides Kashmir, a dispute over water resources is shaping up as a new flashpoint in the unstable region. Last year, India began work on a \$925 million effort to dam the Kishanganga River in the Indus Basin and build a long tunnel to divert water through electricity-generating turbines. But the Kishanganga is a tributary of the Jhelum River, and a 50-year-old treaty gives Pakistan the right to use all of the Jhelum's water. India insists that its "run of the river" hydropower project will not stem flows into the Jhelum. But Pakistan is not convinced—and tempers are flaring.

"India is using water as a weapon against Pakistan," Fateh Ullah Khan Gandapur, former chair of Pakistan's Indus River System Authority, wrote in an editorial last April in Pakistan's largest circulation daily newspaper, *The Dawn*. "If the country faces famine and widespread hunger, ... which the loss of crops will almost certainly generate," he argued, "the resultant situation could lead to a nuclear war at worst, and the conversion of this part of the world into a center for terror activities at best." Pakistan is expected to raise Kishanganga in talks between the two country's foreign ministers next month in Islamabad.

The dispute highlights growing concerns about water security in the Indus River basin, home to more than 200 million people in India and Pakistan and inhabited since the Indus Valley civilization arose 5000 years

ago. The 3200-kilometer-long river supplies water to the world's longest contiguous irrigation system, which covers a 1.1-million-square-kilometer area of the Indus basin. There's hardly a drop to spare. The Indus reaches the Arabian Sea for only about 2 months during the monsoon season; the rest of the year, it dries up in its tracks.

The 1960 Indus Water Treaty gives Pakistan 80% of the approximately 180 billion cubic meters of water per year flowing down the Indus and its tributaries, with the rest allotted to India. But the pact is under siege. India is building a slew of hydroelectric projects that Pakistan officials contend will reduce water flow. Last March, Pakistan raised water security as a concern for the first time in years at high-level talks in New Delhi when Foreign Secretary Salman Bashir accused India of causing "water scarcity" and sought a moratorium on dam building in Indian Kashmir.

Indian officials have said little publicly about the recent water woes. In a speech last April in Karachi, Sharat Sabharwal, India's ambassador to Pakistan, said that "India has never hindered water flows to which Pakistan is entitled, ... and we have no intention of doing so." Any reduced flows were from natural shortages "and not because of any diversion of water by India," Sabharwal said. Indian officials have vowed to press ahead with the dams and suggest that Pakistan is mismanaging its water resources.

Sacred resource. Prayer flags over the Indus River.

The situation is likely to grow tenser. For starters, India and Pakistan are locked in a disagreement over transboundary water data. "The Indian government seems to be very opposed to any data sharing, so it is difficult to develop anything but the most general impression of stream-flow regimes," says Donald Alford, a hydrology consultant based in Billings, Montana. He and others argue that the Indus Water Treaty should be amended to account for anticipated shifts in water resources in response to climate change.

Some experts believe that providing trusted data on water flow might ease tensions and get leaders to focus squarely on a crisis that's expected to grow worse for everyone in the Indus basin. "We need to develop positions on Indus water management that are based on data and not name-calling," says economist Pervaiz Amir, climate change adviser to Pakistan Prime Minister Yousaf Raza Gilani. Raising the stakes, Jamaat-ud-Dawa, a Pakistan-based extremist group linked to Al-Qaeda, has called for a war on India to reclaim Pakistan's "rightful water share."

"I see a looming train wreck on the Indus with disastrous consequences for both countries," warns John Briscoe, a hydrologist at Harvard University.

Sharing the basin

In 1947, when India and Pakistan gained independence and new borders were drawn, the waters of the Indus also needed to be divided. With mediation from the World Bank, the two countries signed the landmark Indus Water Treaty, which gives Pakistan rights in perpetuity over the waters of three west-flowing rivers: the Indus, the Jhelum, and the Chenab. India was allotted rights to the east-flowing Sutlej, Beas, and Ravi rivers. The pact has been a rare point of consensus between two countries that have waged three wars against each other. During the Kargil conflict in 1999, when dialogue broke down between India and Pakistan, expert teams from the Indus Water Commission continued to meet.

The first major dispute mediated under the Indus treaty occurred in 2005, when Pakistan claimed that the 450-megawatt Baglihar hydroelectric project being built on the Chenab River would impede water flows into Pakistan. The treaty permits India to use waters of the west-flowing rivers on its territory in a "nonconsumptive" manner, including for hydropower. For the first time, the Indus Water Commission called in an inde-

pendent expert, Raymond Lafitte, a civil engineer at École Polytechnique Fédérale at Lausanne, Switzerland, to help mediate. Lafitte ruled that India had a right to build the dam but recommended design changes that would reduce the reservoir's storage capacity—changes that India duly implemented.

But suspicion lingers, and India of late has gone on a dam-building spree, with at least 33 hydroelectric projects with a total installed capacity of about 3000 megawatts now under construction on the western rivers. “If Pakistan and India had normal, trustful relations, there would be a mutually verified monitoring process which would assure that there is no change in the flows going into Pakistan,” says Briscoe.

This trust deficit reared in August 2008, when India began to fill the Baglihar reservoir. The timing could not have been worse—“exactly when it would impose maximum harm on farmers in downstream Pakistan,” Briscoe says. The Chenab almost dried up, and crops withered. India's reading of the Indus treaty, Sabharwal says, is that it mandates the timing of impoundment of any Chenab dam. Indeed, a clause states that “initial filling ... on the Chenab can be done between June 21 and August 31.”

Pakistan officials are now taking aim at the 330-megawatt Kishanganga project—in particular, the dam's dead storage capacity for accumulating silt that, they contend, will reduce flows to the Jhelum. Indian Prime Minister Manmohan Singh and Pakistani Prime Minister Yousaf Raza Gilani are reported to have discussed Kishanganga at a summit in Bhutan last April; no details were released.

Shaky forecast

Political dialogue is not the only information held close to the vest. Pakistan officials accuse India of providing low-quality, incomplete, and often tardy data from its Indus basin water gauges. Some experts agree: “The Indian government protects [water-flow data] like a state secret,” says Daanish Mustafa, a geographer at King's College London. In a statement last March, India's Foreign Office said that data are shared with Pakistan every 3 months, as mandated by the Indus treaty.

To obviate concerns about data sharing, Pakistan's treaty commissioner proposed at a recent meeting in Lahore that the two nations together deploy a satellite-based system for online water data sharing. Indian officials say the government is studying the proposal.

But better data may not immediately solve the problems in Pakistan, where data integrity is a big concern. Squabbles have erupted between western Punjab and the downstream Sindh and Baluchistan provinces. “The three regions are fighting among themselves, as there is a shortage of water and each province doubts the veracity of the data provided by the other,” says Daniyal Hashmi, a civil engineer with the Water and Power Development Authority in Lahore. A 5-year-old, \$430 million project by Pakistan's Indus River Systems Authority to post online water-flow data from all 45 canals, 19 barrages, and the two main dams on the Indus and the Jhelum has failed to ease tensions because provincial officials



Calling for calm. Hameed Ullah Jan Afridi.

believe the data are regularly tampered with, Hashmi says.

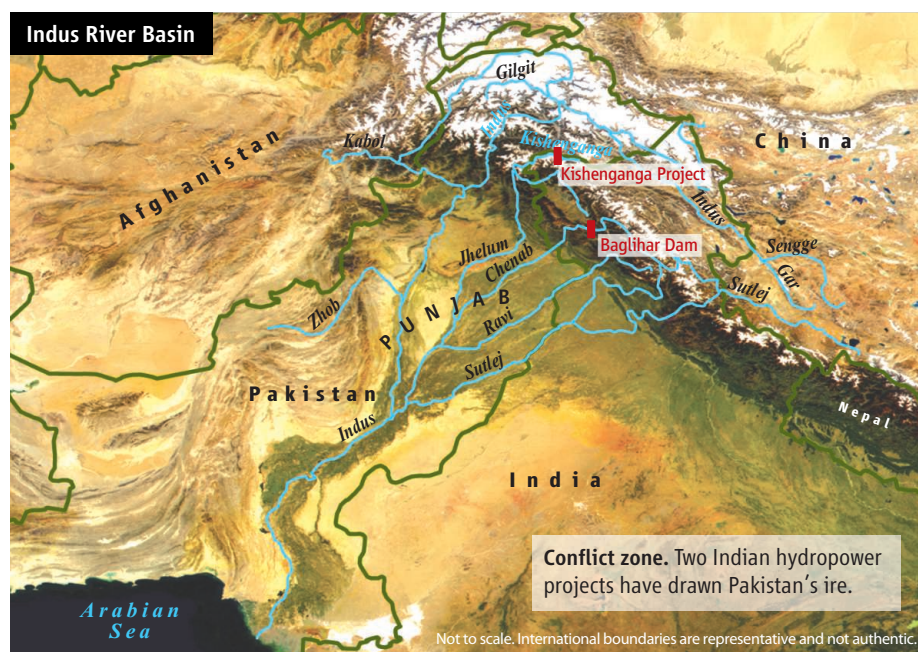
More rational water use might ease concerns. In Pakistan's western Punjab, cropping intensity—the amount of crops grown and the sequence of planting and keeping fields fallow—results in three to four times more vegetation than the Indus basin irrigation system was designed for, says Mustafa. Therefore, he declares, water use in the Indus basin is “unsustainable.” Water

“overuse” in western Punjab is why the Indus dries up 10 months a year, adds Hashmi. The situation is complex, but Punjab's water consumption does reduce flow, says Jeffrey Kargel, a glaciologist at the University of Arizona, Tucson. “There has always been extreme seasonality of flow, but it is also pretty clear that due in part to the extensive use of the water [in western Punjab] and partly due to climate change, this seasonality is becoming more extreme,” he says.

Clouding the outlook is the potential impact of global warming. Himalayan glacier runoff supplies roughly 40% of the Indus's flow. “If one of the possible worst case climate change scenarios were to happen—say, a substantial shutting down of the summer monsoon over the northern part of the subcontinent—then the glaciers would suddenly assume paramount importance. No one, however, has plans or readiness to deal with such a situation,” says Kenneth Hewitt, a glaciologist at Wilfrid Laurier University in Waterloo, Canada. Most experts agree that in the short term, as glaciers melt, Indus basin flows will surge before tapering off decades from now as ice and snow fields disappear. For that reason, the treaty should be amended to cope with anticipated flow change and “climate-change uncertainty,” says John “Jack” Shroder, a geomorphologist at the University of Nebraska, Omaha.

Pakistani officials say they are eager for a peaceful resolution. “Some prophets of doom may foresee a future filled with conflicts,” Environment Minister Hameed Ullah Jan Afridi told *Science*. “But history bears witness to the fact that cooperation, not conflict, is the most logical response to transboundary water management issues.” Better data-sharing may be a first step toward building trust—and time is of the essence. “If we don't do anything now,” says Umesh K. Haritashya, a geologist at the University of Dayton, Ohio, the Indus basin “can become a resource crisis of gigantic magnitude.”

—PALLAVA BAGLA





LETTERS

edited by Jennifer Sills

Scientific Writing, a Case in Point

THE SPECIAL SECTION ON SCIENCE, LANGUAGE, AND LITERACY (23 April, p. 447) addresses the challenge of reading scientific papers. Yet the articles in this section illustrate the very problem they discuss. Consider the long sentences and the number of brackets, parentheses, commas, and dashes in the second paragraph of C. E. Snow's Perspective (p. 450). It is apparent why even scientists only read papers in their own field. Collectively, the sentences in each paper almost need to be diagrammed in order to be understood. I particularly like how P. van den Broek (p. 453) found it necessary to insert two sets of parentheses in the abstract of his paper.

JOANNE O'CONNELL WHITNEY

Department of Clinical Pharmacy, University of California, San Francisco, CA 94101, USA. E-mail: whitneyj@pharmacy.ucsf.edu



very moderate practice, humans can match Ayumu's performance.

In spite of this basic methodological error, the claim of superior spatial working memory in chimpanzees has been widely and uncritically repeated in the popular and scientific media. Propagation of this incorrect idea distracts from more fruitful explorations of chimpanzee memory and undermines ongoing research into human and primate evolution.

PETER COOK* AND MARGARET WILSON

Department of Psychology, University of California-Santa Cruz, Santa Cruz, CA 95064, USA.

*To whom correspondence should be addressed. E-mail: pcook@ucsc.edu

References

1. S. Inoue, T. Matsuzawa, *Curr. Biol.* **17**, 1004 (2007).
2. A. Silberberg, D. Kearns, *Anim. Cognit.* **12**, 405 (2009).

Facing the Facts
on the Public's Beliefs

IN THE NEWS OF THE WEEK STORY "NSF BOARD draws flak for dropping evolution from *Indicators*" (9 April, p. 150), Y. Bhattacharjee reports that the 2010 edition of *Science and Engineering Indicators* excluded data about the public's dismal response to two completely factual statements about human evolution and the big bang theory because reviewers felt the responses "conflated knowledge and beliefs."

Where is the conflation? The statements to which the survey respondents reacted are correct. One either knows that or does not. Would the NSF Science Board see conflation of knowledge and beliefs if 55% of the public responded incorrectly to the statement "Smallpox is caused by a virus" because those respondents believe the disease is caused by demonic possession? Or would the board see those responses for what they truly represent: an outdated and impoverished understanding of the natural world?

When facts conflict with beliefs, it is the beliefs that must give way. The scientific community should not recoil from strong support

of the scientific facts, nor should scientific bodies refrain from sharing data that reveal that much of the public does not understand central facts about the world and the universe in which we live.

JOSEPH D. MCINERNEY¹* AND
MICHAEL J. DOUGHERTY²

¹National Coalition for Health Professional Education in Genetics, Lutherville, MD 21093, USA. ²American Society of Human Genetics, Bethesda, MD 20814, USA.

*To whom correspondence should be addressed. E-mail: jdmcinerney@nchpeg.org

In Practice, Chimp Memory
Study Flawed

IN THE NEWS FOCUS STORY "DID WORKING memory spark creative culture?" (9 April, p. 160), M. Balter reports that "[c]himpanzees are better than humans at some memory tasks," based on work with the chimpanzee Ayumu (*1*). In fact, that study contained a fundamental flaw.

Ayumu received extensive practice on the task; the humans to whom he was compared received none. At least one subsequent study (*2*) shows that, with even

China and India:
Think Outside the Borders

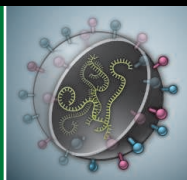
IN THEIR POLICY FORUM "CHINA, INDIA, AND the environment" (19 March, p. 1457), K. S. Bawa and colleagues argue that a bilateral engagement between China and India "will be vital for mitigating biodiversity loss, global warming, and deforestation." Nobody doubts that bilateral cooperation between these two key nations is crucial for resolving such transboundary issues. However, there is considerable doubt as to whether Sino-Indian cooperation is best developed by concentrating on these two nations alone.

China's impact on the environment extends much farther than India's; Beijing's interest in gaining access to Africa's raw materials presents a far more urgent cause for concern (*1*). Therefore, China and India must cooperate beyond the regional level in order to give their efforts a broader, global perspective. For example, if the two join forces with Brazil and Russia—two other nations now rising from underdevelopment—their joint efforts will affect 42% of the world's population. Their plans could encompass environmental



Measuring
the atmosphere

1241



Drug resistance
by flu virus

1243

security strategy as well as cooperation on green technologies and innovation.

Bawa *et al.* proceed from the assumption that the military presence of China and India along their disputed border is damaging the Himalayan ecosystems. However, the authors have overlooked the possibility that, as perverse as it may seem, the fact that the Himalayan region is heavily militarized may actually protect its key habitats and rich biodiversity, because the area is not available for economic exploitation (2). The militaries of the two nations have a negligible environmental footprint in the Himalayas. Both countries use special fuel which must be airlifted to the location, and both are careful to cover the tracks of their deployments, for a variety of strategic reasons.

To mix a sensitive territorial dispute with environmental efforts is a recipe for paralysis. At least for the moment, the Himalayas can wait; the world cannot.

KELVIN S.-H. PEH

Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. E-mail: kelvin.peh@gmail.com

References

1. K. S.-H. Peh, J. Eyal, *Energy Pol.* **38**, 4729 (2010).
2. G. E. Machlis, T. Hanson, *BioScience* **58**, 729 (2008).

Mental Illness Requires a Multidisciplinary Plan

THE POLICY FORUM BY H. AKIL *ET AL.* ("THE future of psychiatric research: Genomes and neural circuits," 26 March, p. 1580) proposed a research project in psychiatry for combining the study of neural circuits and genomics. We do not oppose the project but believe that it is one-sided and in need of a more comprehensive and integrated view.

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 3 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

Akil *et al.* seem to base their proposal on the premise that genetic factors are necessary, if not sufficient, conditions for the alterations in brain structure and function that are conducive to the development of mental illnesses. These assumptions suggest that genes lead to mental illnesses, whereas environmental factors play only a secondary role as triggers on the pathway from genes to the phenotype.

However, the development of the brain and its regulation by gene expression are determined not only by cellular and neural network factors, but also by environmental factors, such as nursing and maternal care (1–4). Gene expression and brain organization are linked to environmental factors by bidirectional pathways, and it is the interactions of all three that lead to psychiatric disorders (5). The extensive study on adoptive children by Tienari *et al.* demonstrates that gene-environment interaction is significant in the development of schizophrenia spectrum disorders (6, 7). Even

psychotherapy has been shown to associate with increase of serotonin-driven brain function (8). We call for a more multidisciplinary view than that represented by Akil *et al.* when planning the future of psychiatric research.

PETTER PORTIN,^{1*} JOHANNES LEHTONEN,²
YRJÖ O. ALANEN³

¹Laboratory of Genetics, Department of Biology, University of Turku, FIN-20014 Turku, Finland. ²Department of Psychiatry, University of Eastern Finland, Kuopio, FIN-70211, Finland. ³Psychiatry, Institute of Clinical Medicine, University of Turku, Turku, FIN-20014, Finland.

*To whom correspondence should be addressed. E-mail: petter.portin@utu.fi

References

1. M. J. Meaney, M. Szyf, *Trends Neurosci.* **28**, 456 (2005).
2. D. A. Lewis, P. Levitt, *Ann. Rev. Neurosci.* **25**, 409 (2002).
3. J. Lehtonen *et al.*, *J. Pediatr.* **132**, 646 (1998).
4. J. Lehtonen *et al.*, *Psychophysiology* **39**, 73 (2002).
5. Y. O. Alanen, *Psychosis* **1**, 156 (2009).
6. P. Tienari *et al.*, *Am. J. Psychiatry* **160**, 1 (2003).
7. P. Tienari *et al.*, *Brit. J. Psychiatry* **184**, 216 (2004).
8. S. M. Lehto *et al.*, *Prog. Neuropsychopharmacol. Biol. Psychiatry* **32**, 229 (2008).

CORRECTIONS AND CLARIFICATIONS

Reports: "Small RNA duplexes function as mobile silencing signals between plant cells" by P. Dunoyer *et al.* (14 May, p. 912). References 19 to 22 are incorrect in the reference list, although they are cited with the correct numbers in the text. Reference 22 by P. Dunoyer *et al.* should be reference 19, and references 19 to 21 should instead be 20 to 22, respectively.



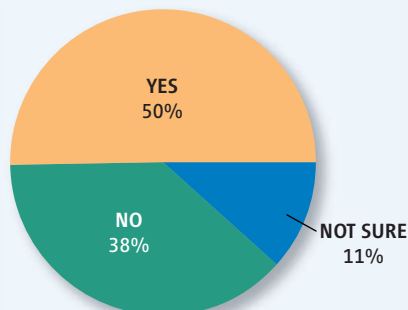
Readers' Poll Results

Unconventional Journals

On 23 April, we asked what you thought about this question:

In general, do the benefits of a journal such as *Medical Hypotheses* outweigh the risks?*

More than 1300 of you responded, from more than 50 countries. Here are the results:



A selection of your thoughts:

"The history of science has shown that the crazy radical idea of a new generation becomes the respected scientific fact of the next. Self censoring of knowledge and ideas has always proved to be a failure."

—commenter David Mayne

"The problem arises when those who are not equipped to analyze some given research just take it for what it is because it is in a journal."

—commenter Peng Liu

*See the poll, and links to the related Letters and News story, at www.sciencemag.org/extra/polls/20100423-1.dtl.

Polling results reflect the votes of those who chose to participate; they do not represent a random sample of the population.

ESSAY REVIEW

The Climate Change Debates

Philip Kitcher

In one of the earliest and most eloquent pleas for open discussion and debate, John Milton wrote:

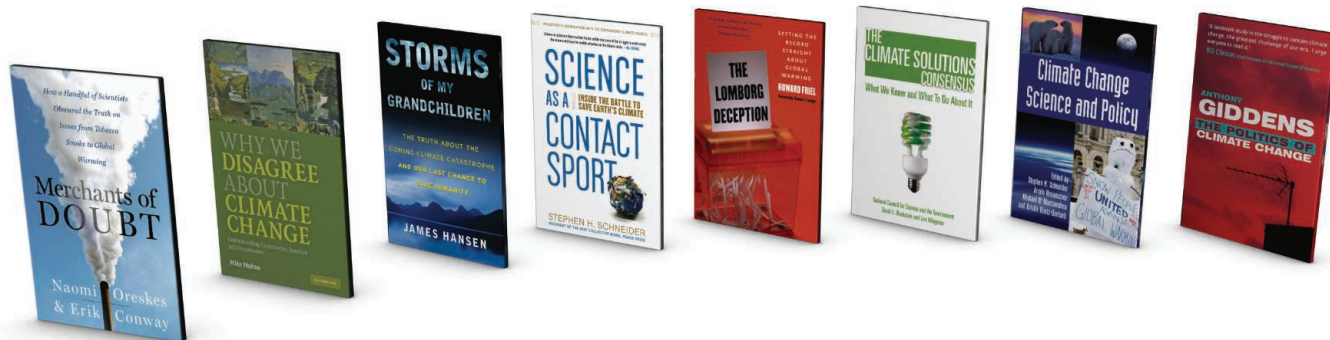
And though all the winds of doctrine were let loose to play upon the earth, so Truth be in the field, we do injuriously, by licensing and prohibiting, to misdoubt her strength. Let her and Falsehood grapple; who ever knew Truth put to the worse, in a free and open encounter. (*I*)

Two centuries after Milton, in the same year in which Charles Darwin published the *Origin*, John Stuart Mill's essay *On Liberty* (2) added further arguments for the free exchange of ideas, suggesting that such exchange is vital for intellectual and social health. Although both Milton and Mill stand behind our current acquiescence in the value of extensive free discussion, both of them

tives on public discussion and policy bears on our own times, although the risks may affect our species as a whole and the stakes may be far higher. For three decades, prominent climate scientists have been warning of the dangerous effects of the continual emission of greenhouse gases into Earth's atmosphere. They have been attempting to identify and to explain just what those effects are likely to be—for ourselves, our children, and our more remote descendants. And they have been urging a variety of measures that might prevent some of the disasters whose possibility they claim to foresee. Yet it is evident that substantial disagreement remains about the consequences for humans and for other species. This is so even in those countries where citizens have largely accepted the conclusions that anthropogenic global warming exists and is likely to raise the average temperature on our planet at least 2°C by the end

voices sometimes muffled. In *Storms of My Grandchildren*, Hansen attempts to combine the story of his own efforts with (yet another) attempt to explain the pertinent parts of climate science as clearly as he can. *Science as a Contact Sport* presents Schneider's insider account of the struggles to understand and moderate human-induced atmospheric changes. Other climate scientists, like Mike Hulme (University of East Anglia), who live in societies where the level of discussion has usually been more informed, are inclined to see matters differently. They hold that continued debate reflects the genuine difficulties of the underlying issues and sometimes explicitly chide their colleagues (as Hulme does in *Why We Disagree About Climate Change*) for a tendency to "apocalyptic" pronouncements. So, in reflections on the debates of the past decades, there opens up a genuine dispute about the role of scientists in influencing public policy, with some urging a stronger voice for expert testimony and others recommending reticence and even quietism.

In part, the differences between Hansen and Schneider, on the one hand, and Hulme, on the other, stem from their concerns with



knew that they were opposing ancient suspicions about the viability of democracy. The political theorists and philosophers of the Greco-Roman world viewed ordinary folk as vulnerable to deception and exploitation. Allowed to determine the direction of the state, the folk would be easily seduced into believing falsehoods aligned with the interests of charismatic leaders, so that the popular voice would enthusiastically clamor for disastrous policies. Better, then, to entrust the ship of state to wise navigators, whose wisdom embraced both depth of understanding and moral integrity.

The contrast between these two perspec-

of the century. In the United States, the state of discussion is less advanced: Denying the reality of human-caused climate change continues to figure as a serious possibility in public debates. And a large fraction of the populace believes that scientists' warnings about the impact of any increases in global temperatures are exaggerated.

For those who play the role of Cassandra in this drama, such as climatologists James Hansen (NASA Goddard Institute for Space Studies) and Stephen Schneider (Stanford University), a 30-year effort to alert policymakers, politicians, and the public to what they perceive as significant dangers can only be seen as frustrating. They have been moved to write books, accessible to a general readership, that will record the ways in which their warnings have been ignored—and their

rather different controversies. It is useful to differentiate three questions. First is the issue of whether human activities, specifically actions that increase the emission of greenhouse gases, are contributing to a significant average warming of Earth. (As all the expert authors point out very clearly, there is no suggestion that the temperature of every region will rise during the next decades.) Second are questions about the probabilities with which various phenomena (complete melting of ice sheets, for example) will occur and about their consequences for human beings and other species. Third are considerations about what might be done to halt (or even reverse) the warming and to limit the damaging consequences. Hulme emphasizes the complexity of the third set of issues. He notes how they are intertwined with difficul-

The reviewer is in the Department of Philosophy, Columbia University, New York, NY 10027, USA. E-mail: psk16@columbia.edu

ties about understanding economic trends and changes, about global justice, about the values assigned to things that are hard to assess in economic terms (ecosystems, the continuation of particular forms of human social life), about practical geopolitics, and even about religious perspectives. Focusing on this intricate web of problems, he elaborates an extensive case for the naturalness of continued disagreement.

For Hansen and Schneider, however, the first two questions are primary (although Hansen ventures some proposals about the third as well). Both contributed to repeated attempts to persuade successive American administrations of the existence and importance of anthropogenic global warming, and Schneider participated in lengthy discussions during the preparation of Intergovernmental Panel on Climate Change (IPCC) reports—discussions in which voices representing political interests seem to have forced compromising the eventual presentation of the pertinent scientific ideas. Their experiences incline them to emphasize the importance of expert judgment, effectively renewing the ancient worries about the dangers of democracy. Both believe that genuine democratic participation in the issues can only begin when citizens are in a position to understand what kinds of policies promote their interests. To achieve that requires a far clearer and unmistakable communication of the consensus views of climate scientists, with respect to the existence of anthropogenic global warming and to the chances of various effects, than has hitherto been available. In his choice of title, Hansen implicitly questions the frequent assumption that effects on future generations are subject to some “deep discount.” He explicitly notes that people’s common concern for the fates of their children and grandchildren provides a shared starting point for responding to the changes that might threaten them. Consequently, if citizens are to be able to express their views about things that matter most to them, they need informed views about the planet on which their descendants will live. Serious democracy requires reliance on expert opinion.

It is all too easy to be beguiled by an opposite thought: that democracy demands that there be extensive public discussion, even on technical matters, discussion in which all participants operate as equals. Those in the grip of this idea will view Hansen and Schneider as hysterical and arrogant people who aim to short-circuit the proper air-

ing of alternative views. (Although sympathetic critics might also ponder the fact that these two eminent scientists have been rebutting the same “alternatives” for decades). Perhaps continued discussion could be tolerated, were there no urgency about the issue under debate. If they saw no compulsion to act soon—and if they were convinced that the fight were fair—Hansen and Schneider might share Milton’s confidence that truth would ultimately emerge as victor. Yet the stories they tell in their gripping narratives reveal all too many points at which messages have been distorted and suppressed

because of the short-term interests of economic and political agents. They also demonstrate many ways in which the arena of public discussion has been set up to block the widespread acceptance of conclusions based on an increasing body of evidence.

The insiders’ stories of ways in which crucial information has effectively been withheld from voters, particularly in the United States, should give us pause about the functioning of our democracy. Even more powerful is the account provided by two outstanding historians who have reviewed a sequence of controversies around topics of public concern. In their fascinating and important study, *Merchants of Doubt*, Naomi Oreskes and Erik M. Conway offer convincing evidence for a surprising and disturbing thesis. Opposition to scientifically well-supported claims about the dangers of cigarette smoking, the difficulties of the Strategic Defense Initiative (“Star Wars”), the effects of acid rain, the existence of the ozone hole, the problems caused by secondhand smoke, and—ultimately—the existence of anthropogenic climate change was used in “the service of political goals and commercial interests” to obstruct the transmission to the American public of important information. Amazingly, the same small cadre of obfuscators figured in all these episodes.

Oreskes (University of California, San Diego) and Conway (NASA’s Jet Propulsion Laboratory) painstakingly trace the ways in which a few scientists, with strong ties to particular industries and with conservative political connections, have played a disproportionate role in debates about controversial questions, influencing policy-makers and the general public alike. Typically, these scientists have obtained their stature in fields other than those most pertinent to the debated question. Yet they have been able to cast enough doubt on the consensus views arrived at by scientists within the relevant disciplines to

Online

sciencemag.org

To join a discussion of the issues raised here, go to <http://tiny.cc/clichng>

Merchants of Doubt

How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming

by **Naomi Oreskes and Erik M. Conway**

Bloomsbury, New York, 2010. 365 pp. \$27, £25. ISBN 9781596916104.

Why We Disagree About Climate Change

Understanding Controversy, Inaction and Opportunity

by **Mike Hulme**

Cambridge University Press, Cambridge, 2009. 432 pp. \$80, £45. ISBN 9780521898690. Paper, \$29.99, £15.99. ISBN 9780521727327.

Storms of My Grandchildren

The Truth About the Coming Climate Catastrophe and Our Last Chance to Save Humanity

by **James Hansen**

Bloomsbury, New York, 2009. 320 pp. \$25, £18. ISBN 9781608192007.

Science as a Contact Sport

Inside the Battle to Save Earth’s Climate

by **Stephen H. Schneider**

National Geographic, Washington, DC, 2009. 303 pp. \$28, C\$35, £16.99. ISBN 9781426205408.

The Lomborg Deception

Setting the Record Straight About Global Warming

by **Howard Friel**

Yale University Press, New Haven, CT, 2010. 270 pp. \$28, £18.99. ISBN 9780300161038.

The Climate Solutions Consensus

by **David E. Blockstein and Leo Wiegman**

Island Press, Washington, DC, 2010. 328 pp. \$50, £31. ISBN 9781597266369. Paper, \$30, £18.99. ISBN 9781597266741.

Climate Change Science and Policy

Stephen H. Schneider, Armin Rosencranz, Michael D. Mastrandrea, and Kristin Kuntz-Duriseti, Eds.

Island Press, Washington, DC, 2010. 542 pp. \$95, £59. ISBN 9781597265669. Paper, \$49.50, £37. ISBN 9781597265676.

The Politics of Climate Change

by **Anthony Giddens**

Polity, Cambridge, 2009. 272 pp. \$69.95, £55. ISBN 9780745646923. Paper, \$22.95, £12.99. ISBN 9780745646930.

delay, often for a substantial period, widespread public acceptance of consequential hypotheses. They have used their stature in whatever areas of science they originally distinguished themselves to pose as experts who express an “alternative view” to the genuinely expert conclusions that seem problematic to the industries that support them or that threaten the ideological directions in which their political allies hope to lead.

The extraordinary story of deliberate obfuscation that Oreskes and Conway document begins with the delight of the tobacco companies in recruiting Fred Seitz and with Seitz’s own connections to “scientists in their twilight years who had turned to fields in which they had no training or experience.” It moves through the forging of a network of industrial and political alliances, and the creation of a variety of institutes and think-tanks devoted to challenging various forms of expert consensus, to a brilliant chapter in which the authors analyze the reasons why, as of 2009, a significant percentage of Americans (43%) continued to dissent from the minimal claim that there is “solid evidence the Earth is warming.” As Oreskes and Conway conclude:

There are many reasons why the United States has failed to act on global warming, but at least one is the confusion raised by Bill Nierenberg, Fred Seitz, and Fred Singer.

This apparently harsh claim is thoroughly justified through a powerful dissection of the ways in which prominent climate scientists, such as Roger Revelle and Ben Santer, were exploited or viciously attacked in the press.

None of this would have been possible without a web of connections among aging scientists, conservative politicians, and executives of companies (particularly those involved in fossil fuels) with a short-term economic interest in denying the impact of the emission of carbon into the atmosphere. But it also could not have produced the broad public skepticism about climate change without help from the media. As Oreskes and Conway point out, “balanced coverage” has become the norm in the dissemination of scientific information. Pitting adversaries against one another for a few minutes has proven an appealing strategy for television news programs to pursue in attracting and retaining viewers. Nor is the idea of “fair and balanced” coverage, in which the viewer (or reader) is allowed to decide, confined to Fox News. Competing “experts” have become common on almost all American radio and television programs, the Internet is awash in adversarial exchanges among those who claim to know, and newspapers, too, “sell” science by framing it as a sport (preferably as much of a contact sport as possible). Oreskes and Conway identify the ways in which the *Washington Times* and the *Wall Street Journal* have nourished the public sense that anthropogenic climate change is a matter of dispute, how they have given disproportionately large space to articles and opinion pieces from the “merchants of doubt,” and how they have sometimes censored the attempts of serious climate scientists to set the record straight. Even the *New York Times*, the American newspaper that takes science reporting most seriously, typically “markets” scientific research by imposing a narrative based on competition among dissenting scientists.

Media contributions to public confusion—what Schneider labels “mediarology”—are elaborated in a number of these books. There is a serious question as to whether American science journalists have conspicuously failed to discharge what might have seemed their central function: to enlighten the public about topics of concern, in areas where an expert consensus has been reached. Howard Friel’s *The Lomborg Deception* offers a careful analysis of the ways in which the “skeptical environmentalist,” Bjørn Lomborg, has selectively used (and sometimes distorted) the available evidence. Friel (an independent scholar whose previous books have critiqued the foreign and Middle East coverage of the *New York Times*) shows how Lomborg’s claims and his status as an expert were uncritically accepted. Apparently, the idea of framing environmental science in terms of a duel between rival “expert perspectives” was too seductive to resist.

For half a century, since the pioneering work of Thomas Kuhn (3), scholars who study the resolution of major scientific debates have understood how complex and difficult judgments about the probative value of data or the significance of unresolved problems can be. The major transitions in the history of the sciences, from the 16th and 17th centuries to the present, have involved intricate debates among competing research programs, among well-informed scientists who gave different weight to particular sorts of evidence. It is an absurd fantasy to believe that citizens who have scant backgrounds in the pertinent field can make responsible decisions about complex technical matters, on the basis of a few five-minute exchanges



CREDIT: STEVE COLE/GETTY IMAGES

among more-or-less articulate speakers or a small number of articles outlining alternative points of view. Democratic ideals have their place in the conduct of inquiry, for it is arguable that there should be more communication between scientists and outsiders in the construction of research agendas, in the discussion of standards of acceptable risk, and in the articulation of policies based on scientific consensus. Genuine democracy, however, requires a division of labor, in which particular groups are charged with the responsibility of resolving questions that bear on the interests of individuals and societies. Other groups, those covering such questions in the media, have the duty to convey the results so that citizens can cast their votes as an enlightened expression of freedom, justifiably aimed at the outcomes for which they hope. Staging a brief disagreement between speakers with supposedly equal credentials, especially when it is not disclosed that one of them is answering to the economic aspirations of a very small segment of the society, is a cynical abnegation of that duty.

Because it is so thorough in disclosing how major policy decisions have been delayed or distorted, *Merchants of Doubt* deserves a wide readership. It is tempting to require that all those engaged in the business of conveying scientific information to the general public should read it. And that science journalists should abandon the obfuscating practice of presenting alternatives with inferior justification as if they were on a par with the scientific consensus.

* * *

Even if American public opinion were reformed overnight, so that virtually all citizens were convinced that anthropogenic global warming is likely to raise the average temperature of the planet by at least 2°C, that would be only the beginning. Beyond that minimal acceptance lie the difficult issues of deciding just what the consequences of a warmer planet will be and what can be done about them. Here, too, denial can easily be induced. Those who want to resist regulatory actions contend that the difficulties that are likely to arise for our descendants have been greatly exaggerated, that whatever problems arise will be addressed by people in a better economic position than we are today, that human beings have shown an admirable ability to adapt to changing environments, and so on and on and on. In countries that have long taken anthropogenic climate change as a settled question, agreeing on the expected consequences and the appropriate response has not

proved easy. American discussions are likely to be haunted by the long denial, so that suspicions about alarmism linger. As psychologists have repeatedly discovered, those who are misinformed and later corrected often lapse into versions of their original error.

Scientists who believe that there are grave consequences for Earth and its future inhabitants face a difficult dilemma. They can talk in probabilistic terms—typically very imprecise probabilistic terms—about possible scenarios. If those potential futures are to be made vivid in ways that might engage citizens and inspire them to action, then the scenarios need to be given in some detail. Yet, as they become more specific, the precision about probabilities goes down, even to the extent that it is only responsible to declare that some outcome lies within the range of possibilities. Occasionally, those who raise the alarm are more definite. If the Arctic ice (including the Greenland ice sheet) melts, polar bears will lose their habitat and the species will go extinct; if sea levels rise in the most probable ways, low-lying islands (and many coastal areas, such as the Ganges delta) will be submerged. Outcomes like these are often met with an uncomfortable shrug. They are to be regretted, of course, but if avoiding them really requires a serious modification of civilized life, then it seems better to adapt: relocate some polar bears to artificially cooled preserves; transport the unfortunate flood victims to higher ground.

Concentration on scenarios that can be presented in detail and also justified as likely entails a serious cost. For it encourages a public perception that these are the only outcomes the Cassandras of climate science fear. A stereotype easily follows. The movement toward action derives from an ideology, one centered in a dislike of competitive market capitalism, a fondness for regulation, a tendency to give priority to the needs of the poor, and an overemphasis on environmental conservation. Global warming is a device used by Birkenstock-wearing, tree-hugging, business-hating liberal intellectuals for advancing their political aims.

“Ideology” is a word that appears relatively frequently in Hulme’s *Why We Disagree About Climate Change* (although he never explains what he means by it). A climatologist who has devoted some serious time to studying history and social studies of science, Hulme aims to offer a broader perspective on the debates that arise once the initial question of the reality of human-caused global warming has been settled. His book is valuable for its diagnosis of the many different levels at which disagreement can arise and the vari-

ety of political stances and value judgments that can incline people to divergent conclusions about what is likely to happen and what might be done. In delineating that diversity, he moves the discussion beyond any appeal to polarized stereotypes: on this side, the captains of industry, their tools, and their dupes; on the other, the flower children in sandals.

Yet Hulme’s book invites misreading. His immersion in the language of various domains of social studies leads him to write as if the theoretical conceptions he deploys in classifying various positions were as reliably grounded as the scientific findings he so clearly and concisely explains. Sometimes, there is even a fashionable indulgence in skeptical distancing, the use of inverted commas (scare quotes) to raise a knowing eyebrow. He announces, for example, that he will tell the story of “how we ‘discovered’ that physical climates could change,” before going on to give a lucid account of how the discovery (real discovery) occurred. In a similar vein, he tells us that the “‘post-normal’ character of climate change” requires a wider range of expert voices, that scientists must concede ground to “other ways of knowing,” and that climate change can become “a mirror into which we can look and see exposed both our individual selves and our collective societies.” The concerned environmentalist who presses on through Hulme’s discussions of the “opportunities” provided for “us” by climate change may eventually give up when he tells his readers to “change our position and examine climate change as an idea of the imagination rather than as a problem to be solved.” Tell it to the Maldives!

That response, however, is too impatient. Hulme’s ideas are more subtle than the (often maladroit) jargon in which he expresses them. If his book more explicitly differentiated areas in which particular groups of people might have greater authority, it would be possible to recognize the value of his diagnoses of the difficulties that attend debates about climate change without supposing that he is advocating the narcissistic quietism his words often suggest. He could accept, for example, the judgment common to Hansen, Schneider, and Oreskes and Conway: that conclusions about the reality of anthropogenic climate change and about the risks that attend some scenarios for the future are matters that can be—and have been—authoritatively decided by a scientific community to which he himself belongs. He should then agree with the implication that, in this domain, it would be foolish to introduce “other ways of knowing.” Hulme could reasonably suppose that the public becomes properly engaged at the

moment when risks have been specified—to the extent that they can be specified—and that citizens' judgments are crucial to decisions about what risks count as acceptable. He could emphasize, as he comments in one of his best discussions, that any decision as to whether a possible future can be tolerated (or even welcomed) should be informed by economic considerations, even though ethical values are crucial to any serious assessment. Finally, his apparently passive recommendation to see ourselves in "the mirror" of climate change—like his Kennedyesque injunction to ask "what climate change can do for us"—can be interpreted more sympathetically as a call for a more systematic investigation of the global challenges that confront us today and those that our descendants will face, one that formulates strategies for safeguarding the future without sacrificing the interests of those currently living.

To make progress on these issues, there will be a need for generally accessible accounts of the likely impact that various levels of global warming will produce. Both Hansen and Schneider describe potential futures, with Schneider being particularly insistent on the important point that scientists owe the public a specification of probabilities (to the extent that that is possible). Two other recent books—*The Climate Solutions Consensus* from the National Council for Science and the Environment and *Climate Change Science and Policy* (for which Schneider served as one of its editors)—offer some helpful and relatively nontechnical information for concerned citizens. The organization of *Climate Change Science and Policy* is particularly valuable, because of the volume's focus on specific types of changes that would affect the lives of future people. It breaks free of the stereotypical concerns about marooned polar bears and dispossessed islanders to emphasize facts about rising sea levels and melting glaciers that are not sufficiently appreciated. Thus Peter Gleick's chapter on water concisely identifies the likely disruption of water supplies and the serious chances of flood-induced pollution. Similarly, Kristie Ebi delivers a useful summary of a variety of ways in which our descendants will probably be more vulnerable to infectious diseases and respiratory conditions. (Although she omits concerns about the possible effects of environmental change on the evolution of disease vectors and cross-species transmission—perhaps because, in assessing these events, the chances are unspecifiable.)

Even though discussions of the predicaments people will face in the future do not exhaust the relevant considerations for decid-

ing what actions we should take now, it is wise to bring them to the fore. Citizens need to understand the challenges with respect to shelter, food supply, water supply, and disease that are likely to arise for their descendants. Hansen's clear perception that an overwhelming majority of the world's population can share a concern about the kinds of lives that will be available to their children and grandchildren is echoed in the decision by the distinguished social theorist Anthony Giddens (London School of Economics) to ground his recommendations in the thesis that "objects in nature can only have value through us" (4). Although some environmentalists would demur, Giddens's approach in *The Politics of Climate Change* has the advantage of increasing the chances for consensus. Like Hulme, he is much concerned to recognize the connections among global problems, insisting, from the beginning, that the challenges of responding to climate change and of meeting the energy needs of the human population must be faced in tandem. He differs from Hulme in not attempting any wide survey of sources of disagreement, and, as readers of his previous works might expect, he is lucid and precise in outlining potential courses of social action. If his book, conceived as a guide for the perplexed citizen, has a flaw, that lies in the breadth and number of the ideas he explores. Those ideas are offered in response to threats he views as profoundly serious:

It will be a colossal task to turn around a society whose whole way of life is constructed around mobility and a 'natural right' to consume energy in a profligate way. Yet it isn't as hopeless an endeavour as it looks.

* * *

All the books reviewed here were written before climate change deniers exultantly exposed the mistakes made by the IPCC in announcing the imminent demise of the Himalayan glaciers and the "conspiratorial e-mails" from the East Anglia climate center. In the wake of these "important revelations," the merchants of doubt were back in business. In December 2009, *Reuters* published a discussion by Singer in which he claimed that the IPCC report was based on "distorted raw data" and algorithms that were not shared with other scientists (5). Few readers of Singer's presentation, or those given by other long-standing climate-change deniers, learned that there is significant independent evidence for Himalayan glacier melt, although not as rapid as the erroneous sentence implied. Probably

still fewer understood that the competitive-cooperative interactions among scientists often involve unguarded remarks about the work of rival "teams," and that references to "tricks" frequently advert to strategies for simplifying complicated mathematical problems or (as in this case) graphical methods of presenting a conclusion perspicuously, rather than to stratagems for deceiving the public. Captured by a naïve and oversimplified image of what "objective science" is like, it is easy for citizens to reject claims of scientific authority when they discover that scientific work is carried out by human beings.

These revelations probably retarded any serious American consensus even on the minimal judgment that is the preliminary to the longer and more difficult debate. Meanwhile, the disappointment at Copenhagen can be seen as evidence that the world is lapsing into a state of resignation. The emissions, of course, do not take a break from the hard decisions.

Nevertheless, there are grounds for the hope expressed by Giddens. Among them is the fact that serious scholars from a variety of crucial disciplines have written valuable books on which future deliberations can build. Those deliberations will require a new synthesis that involves scientists, social scientists, historians—and others, too. It is an embarrassment (at least for me) that philosophers have not contributed more to this necessary conversation. We might clarify some of the methodological issues—for instance, those concerning the variety of risks involved in model-building. Perhaps more important, we could use recent ethical work on responsibilities to future generations and to distant people to articulate a detailed ethical framework that might help a planet's worth of policy-makers find their way to consensus. With luck, a broader group of dedicated scholars may be galvanized by the books discussed here, so that the potential disasters Hansen and Schneider have been warning us about for 30 years will be averted. Perhaps, in the end, truth—and wisdom—will prevail.

References and Notes

1. J. Milton, *Areopagitica* (London, 1644); www.gutenberg.org/etext/608.
2. J. S. Mill, *On Liberty* (London, 1859); http://ebooks.adelaide.edu.au/m/mill/john_stuart/m645o/.
3. T. S. Kuhn, *The Structure of Scientific Revolutions* (Univ. Chicago Press, Chicago, 1962).
4. Giddens derives this view about values from the political philosopher Robert Goodin [see, for example, (6)].
5. <http://blogs.reuters.com/great-debate/tag/climate-change-conference/>.
6. R. Goodin, *Green Political Theory* (Polity, Cambridge, 1992).

10.1126/science.1189312

SUSTAINABILITY

Closing Loopholes: Getting Illegal Fishing Under Control

Stefan Flothmann,¹ Kristin von Kistowski,^{1*} Emily Dolan,² Elsa Lee,¹ Frank Meere,¹ Gunnar Albu³

The current system of port state control lacks transparency, accountability, and the global reach to punish fishers who are illegally emptying our oceans.

Decreasing numbers of fish caught in global fisheries, overcapacity of fishing fleets (1–5), and rising demand for fish (6) heighten the negative impacts of illegal, unreported, and unregulated (IUU) fishing and make it increasingly widespread and profitable for those involved (7–9). This practice undermines sustainable fisheries management (1, 3, 5, 9), particularly on the high seas (international waters beyond the jurisdiction of coastal states) and in coastal waters of developing countries, and has substantial social and economic ramifications (9, 10). Eighty percent of the world's marine fish stocks are fully or overexploited (11). Illegal and unreported fishing alone accounts for catches worth as much as \$23.5 billion annually; this represents an estimated 11 to 26 million tons of fish, equivalent to about one-fifth of the global reported catch (9). Crucially, the more fish stocks are exploited, the more the proportion of illegal catch appears to increase (9).

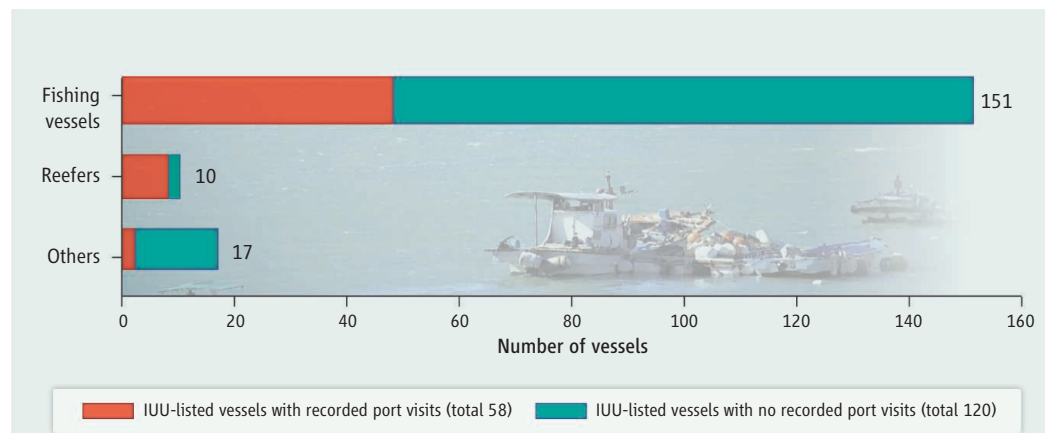
Impacts of fishing activities on the health of fish stocks and their supporting marine ecosystems (3, 12, 13) have spurred new efforts in fisheries management, but their positive effects remain largely unknown (14). We will describe an analysis of the global movements of IUU-listed vessels to evaluate the effectiveness of “port state” measures to combat IUU fishing activities.

Under the United Nations Convention on the Law of the Sea, control of a vessel's activities is the responsibility of the “flag state” to which that vessel is registered (15). When flag states are unable or unwilling to exert such control, other instruments need to be put in place to combat IUU fishing (16). Effective monitoring and control of

fishing activities—requiring identification of IUU fishing and support vessels, sharing of information beyond national borders, and enforcement and sanction—are essential and complementary tools against IUU fishing. International governance bodies have recently turned to port states to help prevent IUU-caught fish from entering international trade and key markets. These port states have the opportunity to reduce substantially the profitability of IUU fishing operations by denying port entry and services to IUU vessels. After agreeing to voluntary port state measures in 2004 (17),

some RFMOs had already adopted port state obligations for member countries. Although some RFMOs require denial of port services or the landing of fish, others go further and require denial of entry into port for vessels known to have been engaged in IUU fishing. In addition, eight RFMOs [supporting online material (SOM), § 1.1] maintain lists of vessels that have been found to carry out or support IUU fishing within the RFMO regulatory area, with the aim of exposing offenders and applying restrictions.

To gauge whether the PSMA could, once in force, lead to a substantial reduction in



Composition and visibility of IUU-listed vessels. Only 33% (58) of the 178 vessels on the combined IUU vessel list show up in any of the port visit sources available to this research. “Reefers” are support vessels that transport fish. “Others” were originally fishing vessels, IUU-listed, but then rebuilt to other functions (e.g., military, tugboats, and petroleum supply vessels).

the United Nations Food and Agriculture Organization (FAO) approved a legally binding Port State Measures Agreement (PSMA) in November 2009 (18). Under the PSMA, port states would be required to close their ports, to prohibit the landing of IUU fish, and to deny port services to vessels that have been engaged in and supporting IUU fishing activities. The question now is whether the PSMA, if signed and ratified by FAO members, will be effective in reducing IUU fishing.

Regional fisheries management organizations (RFMOs), made up of coastal states and fishing nations that work under international agreements to manage specific high seas fisheries or areas, have a central role to play in the fight against IUU fishing. Before the PSMA,

IUU fishing, we evaluated the implementation and effectiveness of port state measures adopted by RFMOs and directed at vessels on the RFMO IUU vessel lists (19). We compiled data on port visits of RFMO IUU-listed vessels from 2004 through 2009, as tracked by publicly available commercial databases (SOM, § 1.2). These data were supplemented with publicly available information from port logs, national fisheries authorities, and RFMOs. Although IUU-listed vessels represent only a small fraction of those operating illegally, they are the only officially recognized IUU vessels and, therefore, provide a basis for evaluating the willingness or capacity of states to implement port state measures and the effectiveness of the current regulations. Our results show that (i) insufficient

¹The Pew Charitable Trusts, The Pew Environment Group, International Ocean Governance, Washington, DC 20004, USA. ²Marine Biology Section, Biology Department, Ghent University, Ghent, B-9000, Belgium. ³Trygg Mat Foundation, Barents Sea Office, Leines, 8285, Norway.

*Author for correspondence. E-mail: kristin@kistowski.de

vessel information, (ii) a lack of compliance by port states, and (iii) the absence of consistent port state measures across regions pose challenges to port state performance.

Insufficient Vessel Information

Port visits of only one-third of the vessels on the IUU vessel lists could be tracked over the six-year period (see the figure on page 1235). It is improbable that this is because the vast majority of IUU-listed vessels ceased operations while listed. It is more likely that many continued to operate unnoticed by not only the commercial databases but also by most national fisheries and port authorities. Many fishing vessels lack unique identifiers, enabling operators of IUU fishing vessels to disguise their identity by renaming vessels or by switching to a different International Radio Call Sign or flag under which to sail. We found that some IUU-listed vessels had changed names up to nine times or flags up to seven times throughout their lifetimes (SOM, § 1.1). The only unique vessel identifier globally available is the International Maritime Organization (IMO) number. However, IMO numbers are not mandatory for fishing vessels, as they are for merchant vessels. Even when vessels had an IMO number, many RFMOs failed to record it on their IUU vessel lists. Without IMO numbers and regularly updated vessel details, IUU vessel lists are rendered largely ineffective when vessels change name or flag.

Lack of Compliance Among Port States

Of the 425 port visits by IUU-listed vessels, 219 were to states that were members of the RFMO that listed the vessel. These vessels should have been identified by port officials and subjected to port state measures. However, port states only fulfilled their obligations in one out of every four cases (fig. S2). Communications with port state officials revealed cases where (i) port officials were not aware of the IUU status of vessels visiting their ports (20); (ii) port officials did not consistently report information on visits by IUU-listed vessels or port state actions to national fisheries authorities; or (iii) measures adopted by RFMOs were not translated into national law, which limited the ability of the authorities to legally execute measures (SOM, § 2.3). Moreover, at the regional level, most RFMOs did not request any information on visits by IUU-listed vessels to the ports of their member states, nor did they consistently assess the compliance of their members with port state measures. They lacked measures to sanction members for failing to meet their obligations.

Regional Focus of Port State Measures

The regional application of current port state measures allows IUU-listed vessels simply to move to other regions when measures are enforced. The North East Atlantic Fisheries Commission (NEAFC) is an RFMO that maintains a comprehensive IUU vessel list (including good record-keeping of IMO numbers), has adopted strict port state measures that extend to fishery support vessels in addition to fishing vessels, and has actively assisted port states in taking action against IUU-listed vessels (21). When NEAFC, in May 2007, included a provision that denied NEAFC IUU-listed vessels entry to the ports of its member states, the proportion of such vessels visiting ports of non-NEAFC-member states doubled (table S1). Although this indicates the desired impact of strengthened port state measures, it also illustrates that, if port state measures remain regional, the problem will shift elsewhere.

Transparent, Accountable, Global

Although port state measures have potential for deterring IUU fishing and eliminating some illegal and unregulated activities in an RFMO's jurisdiction, the PSMA will not have the desired effect unless implemented universally and effectively by port states. Should the PSMA not achieve broad ratification and instead follow the fate of prior international agreements (22, 23), then implementation of port state measures will remain patchy. This will provide continued loopholes for illegal operators.

Accountability requires transparency (24, 25). The inability or unwillingness of port states to share data on port visits and inspection records means that it will be hard to verify whether states that ratified the PSMA go on to implement the agreement effectively. In the absence of a global vessel register and the mandatory use of IMO numbers, or a similar scheme, illegal operators will continue to disguise their vessels easily.

Although port state measures represent only one of a suite of tools to tackle IUU fishing, they can enhance the effectiveness of other monitoring, control, and surveillance mechanisms and market-related measures. Many RFMOs are taking steps in the right direction by establishing or improving port state measures to meet the minimum standards of the PSMA. Nevertheless, most species under the management of the world's RFMOs are declining in biomass. This can be explained in part by ongoing IUU fishing (26) and can only be addressed if all port, flag, and market states make use of a global information sharing system and intensify enforcement.

References and Notes

1. R. Watson, D. Pauly, *Nature* **414**, 534 (2001).
2. R. Hilborn et al., *Annu. Rev. Environ. Resour.* **28**, 359 (2003).
3. D. Pauly et al., *Science* **302**, 1359 (2003).
4. D. Zeller, D. Pauly, *Fish Fish.* **6**, 156 (2005).
5. B. Worm et al., *Science* **325**, 578 (2009).
6. C. L. Delgado, N. S. Wada, M. W. Rosegrant, S. Meijer, M. Ahmed, *Outlook for Fish to 2020: Meeting Global Demand* (International Food Policy Research Institute, Washington, DC, and WorldFish Center, Penang, Malaysia, 2003).
7. High Seas Task Force, *Closing the Net: Stopping Illegal Fishing on the High Seas* (Governments of Australia, Canada, Chile, Namibia, New Zealand, and the United Kingdom; World Wildlife Fund; IUCN; and the Earth Institute at Columbia University, New York, 2006).
8. U. R. Sumaila, J. Alder, H. Keith, *Mar. Policy* **30**, 696 (2006).
9. D. J. Agnew et al., *PLoS ONE* **4**, e4570 (2009).
10. D. J. Agnew, C. T. Barnes, in *Fish Piracy: Combating Illegal, Unreported, and Unregulated Fishing*, Organisation for Economic Cooperation and Development (OECD) Workshop on IUU Fishing Activities, Paris, 19 to 20 April 2004 (OECD Publications, Paris, 2004), pp. 169–200.
11. FAO, *The State of World Fisheries and Aquaculture 2008* (Fisheries and Aquaculture Department, FAO, Rome, 2009).
12. D. Pauly, R. Watson, J. Alder, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **360**, 5 (2005).
13. B. Worm et al., *Science* **314**, 787 (2006).
14. C. Mora et al., *PLoS Biol.* **7**, e1000131 (2009).
15. United Nations Convention on the Law of the Sea, Part XII, § 6: Enforcement, Article 217: Enforcement by Flag States (10 December 1982); www.un.org/Depts/los/index.htm.
16. M. Gianni, W. Simpson, *The Changing Nature of High Seas Fishing: How Flags of Convenience Provide Cover for Illegal, Unreported, and Unregulated Fishing* (Australian Department of Agriculture, Fisheries and Forestry; International Transport Workers' Federation; and World Wildlife Fund International, Gland, Switzerland; 2005).
17. FAO, *Model Scheme on Port State Measures to Combat Illegal, Unreported, and Unregulated Fishing* (FAO, Rome, 2007).
18. FAO, Agreement on Port State Measures to Prevent, Deter and Eliminate Illegal, Unreported and Unregulated Fishing (FAO, Rome, 2009); www.fao.org/Legal/treaties/037s-e.htm.
19. Most RFMOs base fisheries management on lists of authorized vessels, and in most cases, these precede IUU vessel lists. These two lists are not mutually exclusive. Our analysis focused on IUU vessel lists because more meaningful and comparable data were available on port state measures applicable to IUU-listed vessels.
20. For example, vessels were recorded without their IMO number on IUU vessel lists; vessels changed name, flag, and/or call sign; or vessel lists were not kept up to date.
21. For example, the NEAFC secretariat keeps track of IUU-listed vessels and contacts port states approached by such vessels to request action against the vessels.
22. FAO, Agreement to Promote Compliance with International Conservation and Management Measures by Fishing Vessels on the High Seas (FAO, Rome, 1993).
23. T. Pitcher, D. Kalikoski, G. Pramod, K. Short, *Nature* **457**, 658 (2009).
24. P. Langley, *Glob. Soc.* **15**, 73 (2001).
25. A. Gupta, *Glob. Environ. Polit.* **8**, 1 (2008).
26. S. Cullis-Suzuki, D. Pauly, *Mar. Policy*, published online 21 March 2010.
27. Conducted as part of the International Ocean Governance Project of The Pew Environment Group. We thank K. Escobar, A. Fabra, J. Frost, M. Gianni, R. Goldberg, M. Heeb, S. Hulme, F. Kearns, S. Lieberman, B. McIntosh, D. Pauly, and K. Sack for their helpful comments.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1190245/DC1

10.1126/science.1190245

EVOLUTION

Dance Like No One Is Watching, Sing Like No One Is Listening?

Marlene Zuk

I never thought I would be enthusiastic about 24-hour video surveillance. Of course, it can still represent a troubling intrusion of government into the lives of innocent citizens, but as Rodríguez-Muñoz *et al.* (1) demonstrate on page 1269 of this issue, such continual observation of animals can also provide valuable and unusual insights into their ecology and evolution, particularly when directed at a sedentary population of insects. By combining genetic paternity testing with detailed observations facilitated by video technology, the scientists were able to compare laboratory with field findings in an unprecedented fashion, thereby challenging conventional wisdom about sex differences and the determinants of fitness in terms of reproductive success.

Rodríguez-Muñoz *et al.* used the field cricket *Gryllus campestris*, a European species in which males and females spend most of their adult lives in and around burrows in the soil, to test hypotheses about the predictors of male and female reproductive success. A network of 64 motion-sensitive infrared-equipped cameras enabled them to monitor all occupied burrows, snooping on who visited whom until the end of an individual's life. DNA testing enabled determination of parentage for all offspring produced. In effect, the authors were able to perform a study of lifetime reproductive success like those previously done over decades in large, slow-breeding, and difficult-to-observe species like red deer or baboons (2, 3), but in the space of just 2 years. As crickets have been the subject of numerous detailed laboratory studies (4, 5), the field results could be interpreted in the light of findings from much more controlled circumstances.

One of the key findings was that, as theory predicts (6), male reproductive success was more variable than female reproductive success. Although intuitively reasonable, because males are often limited only by the number of females with whom they can



mate, a number potentially far higher than the maximum number of offspring a female can produce and rear, this idea has seldom been tested in the field and virtually never in invertebrates. Interestingly, however, the male crickets' greater variance in reproductive success did not result from a higher mating rate; both sexes mated with multiple partners, and both left more surviving offspring when they had more mates. Rodríguez-Muñoz *et al.* suggest that some males may simply have achieved a high fitness either via success at sperm competition or because their offspring were, in turn, of higher quality. If the latter explanation were to be the case, the study would also provide rare field support for "good genes" models of mate choice (6), which posit that females prefer males with exaggerated secondary sexual characteristics because such males pass on genes for viability to their offspring.

Contrary to theory, or at least to common assumptions about animal behavior and ecology, was the discovery that the vast majority of individuals, whether male or female, did not successfully reproduce at all. Although males are often acknowledged to play a high-stakes, high-risk game with many losers, conventional wisdom has it that virtually all females, even those in relatively poor condition, should be able to eke out at least one or two young; this disparity underlies, for example, the Trivers-Willard effect on sex ratio (7), in which mothers in poor condition are predicted to favor daughters over sons because even low-quality females are expected to be able to reproduce. Yet none of the female *G. campestris* and only a couple of the males had more

Genetic paternity testing and field observations of crickets challenge conventional wisdom about sex differences and fitness.

than 10 surviving offspring and most had none, despite the ability of females to lay scores of eggs (8).

Finally, being the alpha cricket, or at least winning most of one's aggressive encounters, turned out not to be the indicator of fitness it is often supposed to be. Popular and scientific conventional wisdom emphasizes the outcome of fights and the attainment of high social status

as key to reproductive success, as if the animal kingdom is a giant battlefield in which a single victor always reaps the spoils, and those spoils are always the ability to perpetuate one's genes. The truth, of course, is much more complex: Winners of fights over food or other resources do not necessarily mate more; dominance relationships can be fluid; and many species, including crickets, do not live in stable groups with long-standing dominance hierarchies, even though predictable relationships among individuals can be demonstrated in the lab (9, 10). Even among long-lived primates, high rank is not always a clear correlate of reproductive success (11, 12). Without denying the importance of aggression in animal life, Rodríguez-Muñoz *et al.*'s finding that winning fights was unrelated to the number of descendants is a good reminder that conventional wisdom is not always borne out by the facts.

The ability to so closely monitor the lives of animals in the field when they have already been well studied in the lab paves the way for the testing of sometimes hotly contested ideas in behavioral ecology. For example, how costly is female preference, and what price, if any, do females pay for mating with multiple males? Theory can account for female choice of mates with exaggerated secondary sexual characters without any viability benefit, solely via Fisherian runaway processes (13), obviating the need for females to choose males with good genes. These models rely on females not paying much of a cost as they search for mates, but we know little about whether this assumption is realistic. Following the activities of females as they move from

one male to another could help resolve this issue. Being able to go from the field to the laboratory and back is one of the many virtues of studying invertebrates—one can test a hypothesis under exquisitely controlled circumstances, and then ask whether the findings have any relevance in nature. On the other hand, field observations can suffer from too many contributors to variation, so that it is difficult to isolate the key variable and infer cause and effect. Setups like those used by Rodríguez-Muñoz *et al.* could give us the best of both worlds. And maybe

someday they will provide the basis for a new TV show, say “Big Jumping Brother”?

References

1. R. Rodríguez-Muñoz, A. Bretman, J. Slate, C. A. Walling, T. Tregenza, *Science* **328**, 1269 (2010).
2. T. H. Clutton-Brock *et al.*, *Red Deer: Behavior and Ecology of Two Sexes* (Univ. of Chicago Press, Chicago, 1982).
3. www.princeton.edu/~baboon/index.html
4. M. D. Hall, L. F. Bussière, M. Demont, P. I. Ward, R. C. Brooks, *Mol. Ecol.* **19**, 610 (2010).
5. M. L. Thomas, L. W. Simmons, *Behav. Ecol.* **20**, 1118 (2009).
6. M. Andersson, *Sexual Selection* (Princeton Univ. Press, Princeton, NJ, 1994).
7. R. L. Trivers, D. E. Willard, *Science* **179**, 90 (1973).
8. W. Lohr, M. D. Zaretsky, in *Crick Behavior and Neurobiology*, F. Huber *et al.*, Eds. (Cornell Univ. Press, Ithaca, NY, 1990), pp. 114–116.
9. G. Cowlshaw, R. I. M. Dunbar, *Anim. Behav.* **41**, 1045 (1991).
10. N. Kutsukake, C. L. Nunn, *Behav. Ecol. Sociobiol.* **60**, 695 (2006).
11. J. R. de Ruiter, M. Inoue, *Primates* **34**, 553 (1993).
12. J. H. Manson, in *Primates in Perspective*, C. J. Campbell *et al.*, Eds. (Oxford Univ. Press, Oxford, 2007), pp. 447–463.
13. H. Kokko, R. Brooks, J. M. McNamara, A. I. Houston, *Proc. Biol. Sci.* **269**, 1331 (2002).

10.1126/science.1191036

ATMOSPHERIC SCIENCE

Countering the Early Faint Sun

Christopher F. Chyba

In 1972, Carl Sagan and George Mullen noted the earlier discovery that our Sun should have been around 30% less luminous 4.5 billion years ago, and examined the consequences for primordial Earth (1). Main-sequence stars like the Sun burn hydrogen to helium, and as they do the star becomes more luminous. Sagan and Mullen showed that this solar evolution meant that Earth's current 33° of greenhouse warming, primarily from water vapor and carbon dioxide (CO₂), could not have kept Earth's average surface temperatures above freezing prior to around 2 billion years ago (see the figure). The abundant evidence for liquid water at Earth's surface before that date, they argued, constitutes what is now called the “early faint Sun paradox.”

Sagan and Mullen showed that the paradox could be solved by a concentration as low as 10⁻⁵ of the powerful greenhouse gas ammonia (NH₃) in the early atmosphere. But Kuhn and Atreya subsequently demonstrated that ultraviolet (UV) light from the early Sun should have photolyzed this NH₃ in less than 10 years, driving it irreversibly into greenhouse-irrelevant molecular nitrogen, N₂ (2). In a paper he was working on at the time of his death in 1996, Sagan suggested that an early nitrogen-methane (N₂-CH₄) atmosphere would form a self-protecting layer of organic haze, similar to that formed in the atmosphere of Saturn's moon Titan. This haze layer would block UV while letting visible light through, allowing the NH₃ to persist over geological time with reasonable rates of resupply to the atmosphere

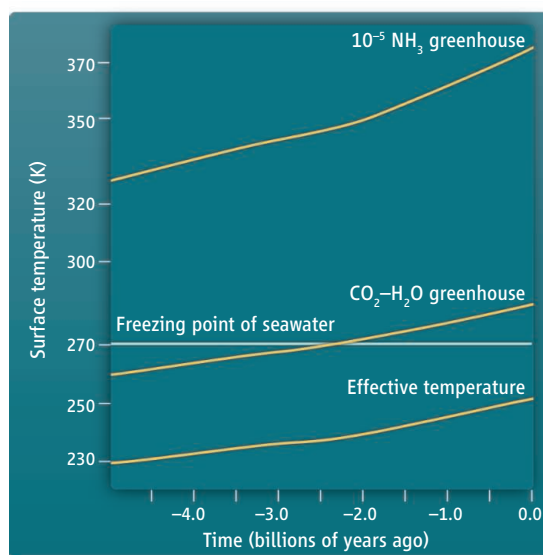
(3). But this idea was attacked with the claim that, to the contrary, the haze would have absorbed too much visible light, so that it would instead have provided an “anti-greenhouse” cooling of the Earth (4, 5). On page 1266 of this issue, Wolf and Toon show that because of the fractal nature of the haze particles, the haze would have been optically thin at visible wavelengths despite being optically thick in the UV (6). Therefore, the early faint Sun problem could have been solved by small amounts of NH₃ after all.

An early-Earth atmosphere with CH₄ and NH₃ is appealing for two reasons. The first is the potential to solve the faint-Sun problem, although the most common reply to

A haze of fractal particles may account for the warm temperatures that led to the wet conditions of the early Earth.

this dilemma—the presence of a thicker CO₂ atmosphere on early Earth—remains as a possible solution (7). The second is the advantages of CH₄-NH₃ atmospheres for one popular scenario for the origin of life. In the “building blocks” view of life's origin, put on an experimental basis by Stanley Miller and Harold Urey in 1952, copious quantities of simple organics such as hydrogen cyanide (HCN) and formaldehyde (H₂CO) are produced in an early CH₄-NH₃ atmosphere, rain out into the oceans, and react to form amino acids (8).

Given these advantages, it is unfortunate that this picture for early Earth faces two major challenges. The first is an argument that this hydrogen-rich (hence reducing) atmosphere should never have formed in the first place. As Earth was accreted from fast-colliding planetesimals, it would have been hot at the time of its formation, and its iron core would have formed almost simultaneously with Earth itself. With the iron largely sequestered in Earth's core, it would not have been available to provide an oxygen sink in Earth's mantle, and carbon and nitrogen would have been outgassed into the atmosphere largely in more oxidized forms (as they are today) rather than as CH₄ and NH₃ (7). But newer models suggest that, even in CO₂-rich early atmospheres, there could have been abundant molecular hydrogen (H₂), recapturing some of the advantages of a reducing atmosphere for the origin of life (9).



Earth's surface temperature. The lines show average Earth surface temperatures through time assuming no greenhouse effect, the current CO₂-H₂O greenhouse (neglecting ice-albedo feedback), and a putative atmosphere with 10⁻⁵ NH₃. Simplified from (1).

Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA. E-mail: cchyba@princeton.edu

The second argument against an early reducing atmosphere has been the UV-destruction objection. The organic-haze model counters this UV photolysis dilemma (3, 6). The remaining question has been whether it also solves, or instead exacerbates, the faint-Sun problem. Wolf and Toon (6) argue that hydrocarbon aerosols are known to form fluffy aggregates of spherical subunits with a fractal structure. These aggregates have optical properties very different from those of the subunits alone. In particular, the fractal particles have ratios of optical depths for UV to visible light of more than 22. Wolf and Toon show that this leads to only an insignificant anti-greenhouse effect, while allowing the underlying NH_3 to provide the greenhouse to counter the early faint Sun. A remaining question is whether the bulk of the NH_3 in the early atmosphere really remains below the UV shield of the organic-haze layer, although this seems plausible (3, 6).

The challenges to the reducing-atmosphere

picture of early Earth have provided impetus to other models for Earth's early environment and the origin of life. Perhaps the most extreme has been a model for early Earth that resembles Jupiter's moon Europa: an ocean of liquid water, heated from within (on Earth, by geothermal heating), but covered by an ice layer too thick to allow photosynthesis (10). Models in which extraterrestrial sources make substantial contributions to Earth's early prebiotic inventory (11), or which reject the building-block approach altogether and examine metabolic reactions that can fix their own carbon as the first steps on the road to life (12), have also flourished.

Important challenges to the building-block approach remain. In particular, polymerization of amino acids into peptides, or nucleotides into RNA, is thermodynamically unfavorable in liquid water. But in fact, many steps in all origin-of-life scenarios remain obscure (13). The origin-of-life and astrobiology scientific communities are addressing these challenges, but

it seems clear that at best a few base-camps have been established in an ascent to a still-invisible summit.

References

1. C. Sagan, G. Mullen, *Science* **177**, 52 (1972).
2. W. R. Kuhn, S. K. Atreya, *Icarus* **37**, 207 (1979).
3. C. Sagan, C. Chyba, *Science* **276**, 1217 (1997).
4. C. P. McKay, R. D. Lorenz, J. I. Lunine, *Icarus* **137**, 56 (1999).
5. A. A. Pavlov, L. L. Brown, J. F. Kasting, *J. Geophys. Res.* **106** (E10), 23,267 (2001).
6. E. T. Wolf, O. B. Toon, *Science* **328**, 1266 (2010).
7. D. Catling, J. F. Kasting, in *Planets and Life*, W. T. Sullivan, J. A. Baross, Eds. (Cambridge Univ. Press, Cambridge, 2007), pp. 91–116.
8. S. L. Miller, *Science* **117**, 528 (1953).
9. F. Tian, O. B. Toon, A. A. Pavlov, H. De Sterck, *Science* **308**, 1014 (2005).
10. J. L. Bada, C. Bigham, S. L. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1248 (1994).
11. C. F. Chyba, K. P. Hand, in *Comets and the Origin and Evolution of Life*, P. J. Thomas, R. D. Hicks, C. F. Chyba, C. P. McKay Eds. (Springer-Verlag, Berlin, ed. 2, 2006), pp. 169–206.
12. G. Wächtershäuser, *Microbiol. Rev.* **52**, 452 (1988).
13. R. Shapiro, in *Planets and Life*, W. T. Sullivan, J. A. Baross, Eds. (Cambridge Univ. Press, Cambridge, 2007).

10.1126/science.1189196

PHYSICS

The Lamb Shift—Yesterday, Today, and Tomorrow

Marlan O. Scully¹ and Anatoly A. Svidzinsky²

The study of the emission and absorption of radiation is the royal road that led Planck to quantum mechanics and Einstein to the concept of the photon. The experiment of Röhlberger *et al.* (1) reported on page 1248 of this issue is another step along the way toward explaining changes in the spontaneous light emission from ensembles of emitters. These effects allow us to probe aspects of quantum electrodynamics (QED) in relatively low-energy nuclear experiments.

To put their experiment in perspective, we first need to compare processes with real particles versus those with virtual particles. Although virtual particles have some of the properties of real ones, the former come into existence and then quickly disappear. Spontaneous emission of real photons occurs when an atom or nucleus jumps from an excited state to a lower-energy state. More intriguing are processes in which the

atom jumps to an excited state and a virtual photon is emitted, followed quickly by the reverse process in which the atom jumps back to the ground state and now absorbs a photon (see the first figure, panel A). These surreal virtual processes have real effects—they can shift the energy levels of emitting atoms, and are called Lamb shifts after Willis Lamb, who first observed them experimentally (2, 3).

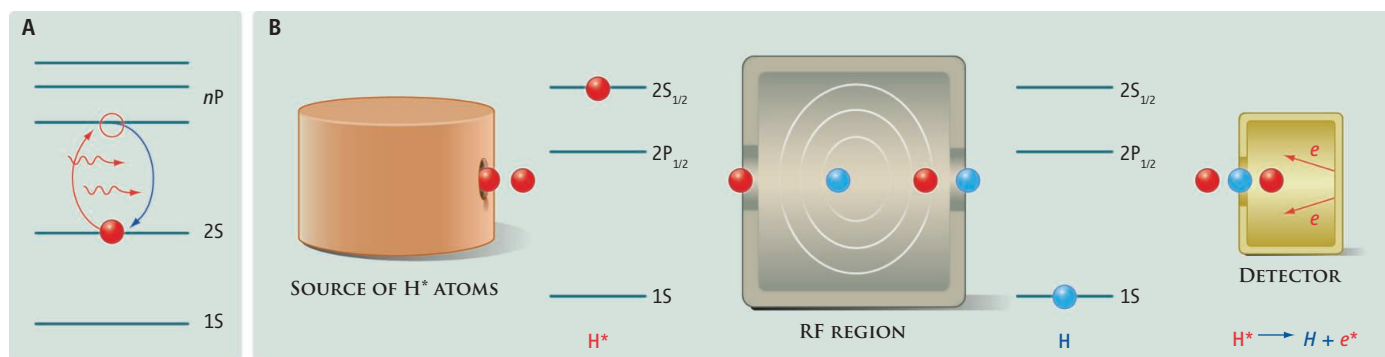
The observation of these shifts garnered Lamb the Nobel Prize because they led to the resolution of fundamental mathematical problems in QED as formulated by Dirac (4). Observable quantities that should be finite diverged to infinity in QED; for example, Oppenheimer's 1930 calculation of the electromagnetic energy of a bound electron—its “self-energy”—led to an infinite level shift (5). No one knew what to make of these infinities. In 1947, Lamb, along with his student Retherford, measured the energy of the electron bound in the hydrogen atom and found a small (not infinite) shift in the 2s atomic energy level (2). A beam of excited hydrogen atoms in the relatively long-lived

Quantum field effects are magnified by collective interactions between many atoms.

$2S_{1/2}$ state was directed onto a detector (see the first figure, panel B). When an atom in the excited state struck the surface, an electron was emitted. The beam was then investigated by means of microwaves, which transferred atoms from the long-lived $2S_{1/2}$ state into the $2P_{1/2}$ level, which then decayed rapidly into the ground state. When the microwave frequency was near the $2S_{1/2} - 2P_{1/2}$ energy, the deexcitation of the atoms led to a drop in the number of emitted electrons. The $2S_{1/2}$ level, which would normally be at the same energy as the $2P_{1/2}$ state, is in fact higher in energy by about 1000 MHz (about 0.004 meV).

The unrealistic infinite level shift of Oppenheimer's calculation was now replaced by the real, finite level shift of Lamb's experiment. Bethe soon presented a new calculation in which he applied ideas of Kramers (6), who had suggested that the infinite self-energy of the bound electron should be “renormalized” by subtracting off the infinite energy of a free electron. Bethe used this renormalization subtraction scheme (7) to calculate a shift of 1040 MHz,

¹Texas A&M University, College Station, TX 77843, USA.
²Princeton University, Princeton, NJ 08544, USA. E-mail: scully@tamu.edu



which was not only finite but also compared well to Lamb's experimental result of 1058 MHz (8). The success of this approach in the hands of Dyson, Feynman, Schwinger, and Tomonaga ushered in modern quantum field theory in very good agreement with experiments (9, 10).

The problem becomes even more fascinating if instead of one atom, N atoms collectively interact via a common electromagnetic field. In 1954, Dicke (11) considered a system of N two-level "spins" confined in a volume whose size is much smaller than the radiation wavelength. Emission from even a weakly excited group of atoms can have a rich structure. If only one of the N atoms is excited, we might expect the radiation rate to be the same as the decay rate of a single isolated atom, γ . However, the Dicke symmetric state of maximum cooperation radiates at a rate proportional to N times γ ; this faster emission rate due to the presence of other atoms is the essence of Dicke superradiance (12). Thus, when an ensemble of atoms emits spontaneously, the emission rate can

be faster than that when they emit singly.

Recent calculations focus on the problem of a single photon stored in a cloud of atoms (it is shared among all the atoms) and then retrieved at a later time. Here, the usual picture of collective emission, which treats each atom as a tiny antenna, no longer applies. Peculiar features of the directional and temporal characteristics of the cooperatively emitted radiation have been predicted (13–15).

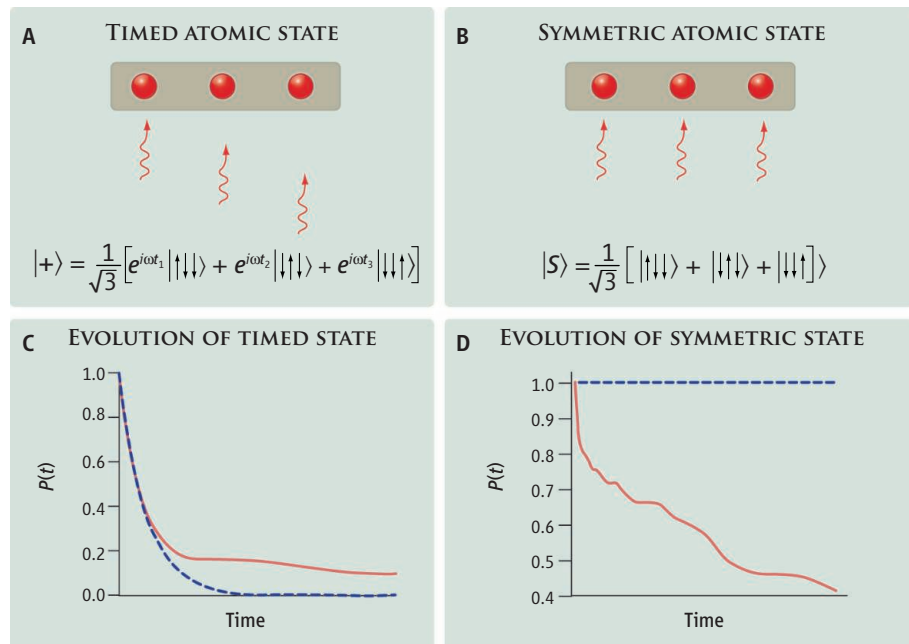
Does the Lamb shift also change in this N -atom system? The Lamb shift is indeed greater when collective emission occurs with many particles (16, 17), and challenging gas-phase studies have verified the collective Lamb shift experimentally (18). However, it is important to recognize that there are obscuring effects (e.g., atom-atom interactions) in the usual Dicke model, which applies when the size of the atomic sample is smaller than the wavelength. To get around the atomic dipole interaction that also shifts energy levels, the atoms should be spread out over a volume that is large rela-

The Lamb shift in unison. (A) The effect of virtual photons on atomic emission is shown for an atom's 2S state. This state is shifted when an electron jumps to a higher n state and emits a virtual photon and returns to the 2S state. (B) A schematic of Lamb's original experiment adapted from (2). A beam of hydrogen atoms prepared in the long-lived 2S_{1/2} excited state passes from the source into a radio-frequency (RF) cavity where 1050-MHz radiation (resonant with the 2S_{1/2} – 2P_{1/2} transition) deexcites the atoms.

tive to the wavelength. It is desirable to have a system in which the atoms are prepared in an optically thin medium, which prevents strong absorption, but are allowed to radiate in an optically thick configuration (13), which enhances their interactions.

Röhlsberger *et al.* created such conditions in their experiment by "sidewise" pumping of emitters (see the second figure, panel A). They studied the collective Lamb shift by exciting an extended (solid-state) array of nuclei into a special superradiant state. These x-ray synchrotron radia-

The Lamb shift collectively. (A) Röhlsberger *et al.* studied a nuclear transition of many iron-57 atoms in a timed Dicke state $|+\rangle$. An arrow up indicates an excited atom and a downward arrow is a deexcited atom. We depict only three atoms and show that the photon (frequency ω) that excites the atoms is delayed. It hits either the first, second, or third atom at specific times t_1 , t_2 , and t_3 . (B) The symmetric Dicke state $|S\rangle$ has no time-dependent phase factors. (C and D) Tomorrow's experiments will likely show that atomic decay of an ensemble of atoms prepared in (C) the timed state $|+\rangle$ or (D) the symmetric state $|S\rangle$ depends on the presence (solid lines) or absence (dashed lines) of virtual photons. $P(t)$ is the probability that atoms are excited as a function of time t . For the timed state, the atoms, although spread out over a large distance relative to the wavelength, undergo superradiant decay. However, at long times, about 10 to 20% of the population is trapped because of the Lamb shift virtual interactions. For the symmetric state, the opposite takes place. The atoms are trapped in the initial subradiant state. However, virtual photons disrupt the highly correlated subradiant state and radiation gradually escapes.



tion experiments involving N nuclei have features in common with collective atomic radiation in the optical domain (19, 20). Synchrotron x-ray photons weakly excite an ensemble of iron-57 nuclei with the proper timing in a slab. Röhlsberger *et al.* measured the temporal evolution of the state decay and the collective Lamb shift by analyzing the energy spectra of the radiation reflected from the sample. The collective Lamb shift almost doubled its value when the particle number doubled.

The experiments of yesterday and today use the atom or nucleus as a laboratory to study virtual processes. As such, the tiny electromagnetic level shifts have given us deep insights into nature. What about the future? Recent calculations show that system dynamics can be strongly influenced by vacuum fluctuations (14, 21); that is, the collective Lamb shift can have a big effect on decay rates. For example, for a large

atomic cloud, a symmetric state $|S\rangle$ (see the second figure, panel B) is trapped. However, virtual transitions couple it to decaying states, which results in a slow decay of the $|S\rangle$ state (see the second figure, panel D). On the other hand, virtual photons only slightly change the evolution of the rapidly decaying states, such as the timed state $|+\rangle$ (see the second figure, panel C). Nevertheless, they change the long time dynamics from the exponential decay into a power-law behavior.

Many other interesting issues are associated with vacuum fluctuations. For example, the collective part of the many-atom Lamb shift is free of infinities (17). In addition, for very large atomic samples, memory effects become important and yield “a new kind of cavity QED” (12). The fascinating effects of virtual photons on N -atom collective emission will be fertile ground for future research.

References and Notes

1. R. Röhlsberger *et al.*, *Science* **328**, 1248 (2010); published online 13 May 2010 (10.1126/science.1187770).
2. W. E. Lamb, R. C. Retherford, *Phys. Rev.* **72**, 241 (1947).
3. In his lectures, Lamb modestly called it the “electromagnetic level shift.”
4. P. A. M. Dirac, *Proc. R. Soc. London Ser. A*, **114**, 243 (1927).
5. J. R. Oppenheimer, *Phys. Rev.* **35**, 461 (1930).
6. H. A. Kramers, *Nuovo Cim.* **15**, 108 (1938).
7. H. A. Bethe, *Phys. Rev.* **72**, 339 (1947).
8. S. Triebwasser *et al.*, *Phys. Rev.* **89**, 98 (1953).
9. A beautiful precision measurement of the 1S Lamb shift of atomic hydrogen and deuterium was pioneered in the study in (10).
10. T. W. Hänsch *et al.*, *Phys. Rev. Lett.* **34**, 307 (1975).
11. R. Dicke, *Phys. Rev.* **93**, 99 (1954).
12. M. O. Scully, A. A. Svidzinsky, *Science* **325**, 1510 (2009).
13. M. O. Scully, *Laser Phys.* **17**, 635 (2007).
14. A. A. Svidzinsky *et al.*, *Phys. Rev. A* **81**, 053821 (2010).
15. S. Prasad, R. J. Glauber, *Phys. Rev. A* **31**, 1583 (1985).
16. R. Friedberg *et al.*, *Phys. Rep.* **C 7**, 101 (1973).
17. M. O. Scully, *Phys. Rev. Lett.* **102**, 143601 (2009).
18. W. R. Garrett *et al.*, *Phys. Rev. Lett.* **64**, 1717 (1990).
19. U. van Bürc, *Hyperfine Interact.* **123–124**, 483 (1999).
20. B. W. Adams, *J. Mod. Opt.* **56**, 1974 (2009).
21. R. Friedberg, *Ann. Phys.* **325**, 345 (2010).

10.1126/science.1190737

ATMOSPHERIC SCIENCE

Top-Down Versus Bottom-Up

Euan Nisbet¹ and Ray Weiss²

Greenhouse gas emissions are currently quantified from statistical data without testing the results against the actual increases of these gases in the atmosphere. This is like dieting without weighing oneself. Data are produced by greenhouse gas emitters of all sizes, from factory or farm to nation, and are quoted to high precision—yet misreporting occurs, whether by simple error, ignorance, or intention. But now scientists on both sides of the Atlantic are arguing that regulation of greenhouse gas emissions can have integrity only if verified by direct atmospheric measurements (1, 2).

Measurement methods have improved remarkably in the past few years, especially with the advent of new optical methods that provide continuous high-precision carbon dioxide (CO₂), methane (CH₄), and nitrous oxide (N₂O) data relatively inexpensively, and with prospects of added stable isotopic data. Modeling methodologies are under development that promise evolution from coarse global-scale understanding to resolution of regional- or country-scale emissions (3).

Currently, carbon-equivalent emissions are assessed by “bottom-up” methods, which aggregate disparate local statistics such as fuel consumption or numbers of cows. Global-scale emissions of some industrial greenhouse gases, as measured by their accumulation in the atmosphere, can disagree with reported bottom-up emissions by factors of two or more (4–7). Atmospheric data also suggest that actual emissions of industrial greenhouse gases tend to be greater than reported.

To carry out “top-down” assessment, that is, using atmospheric understanding to quantify emissions, an approach that integrates several methods is needed. First, the atmosphere must be measured at high spatial and temporal resolution via networks of ground-based stations and aircraft (see the figures). Second, remote sensing is needed, both from satellites to give global coverage and from the ground to calibrate the satellite data. Third, modeling synthesizes the results and assesses budgets. As the data collection network and data interpretation through modeling improve, we can begin to envision their use to test and validate bottom-up inventories.

Can national emissions inventories be verified through direct atmospheric measurements?

Most emissions and many sinks are in the atmospheric boundary layer, the air next to the ground. Atmospheric contents become well mixed worldwide within a few years. However, local composition can vary greatly. For example, the CO₂ atmospheric mixing ratio, currently about 388 parts per million in the global background, can locally be significantly reduced by the springtime growth spurt in a deciduous forest or rise dramatically in urban winter rush hours. Long-term local measurements carry much information about sources and sinks.



Measuring the atmosphere via aircraft. NASA's Global Hawk robotic aircraft—with a range of 20,000 km, an altitude capability above 18 km, and 30 hours of endurance—is being developed in collaboration with NOAA to measure and sample atmospheric greenhouse gases and other climate-relevant parameters.

¹Department of Earth Sciences, Royal Holloway, University of London, Egham TW20 0EX, UK. ²Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093-0244, USA. E-mail: e.nisbet@es.rhul.ac.uk; rfweiss@ucsd.edu



A ground-based station. The Sphinx high-altitude (3571 m) research station at the Jungfrauoch in the Bernese Alps, Switzerland, as seen from the MetAir atmospheric research aircraft. High-frequency measurements of greenhouse gases at this station are combined with similar measurements at other stations in Europe and around the globe in top-down modeling of regional greenhouse gas emissions.

To obtain these data, air collected from towers feeds instruments that continuously record regional greenhouse gas variations. On remote coasts and islands, the marine background is monitored to map the global state of the atmosphere, continuing the work begun by Charles David Keeling more than 50 years ago (8).

Multigas in situ records can be used to track seasonal and diurnal emission patterns and to test emissions declarations for internal consistency (9). Single stations too are valuable (10). Fluxes can be quantified locally by comparison with otherwise-known emissions, such as the radon isotope ^{222}Rn . If the CO_2/CH_4 ratios in the local air do not match inventory estimates for a major urban area, the inventory should be reexamined. Isotopes are powerful identifiers of sources, especially if measurements are coupled with models of atmospheric transport and mixing. Fossil and nonfossil sources of CO_2 can be differentiated by measuring ^{14}C in atmospheric CO_2 (11). The ratios of ^{13}C to ^{12}C in CH_4 can differentiate between wetland sources and clathrate and gas field sources.

Aircraft studies are important, for example, in showing that CO_2 uptake by tropical ecosystems is globally important (12). Satellites add the wider dimension and will be

especially valuable in tropical regions where long-term in situ data are scarce. Currently, satellites such as the aging European Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) instrument and Japan's Greenhouse Gases Observing Satellite (GOSAT) are mapping carbon gases by means of reflected sunlight, and the United States is planning to replace its lost Orbiting Carbon Observatory (OCO) satellite. Quantitative verification of emissions inventories is still beyond the present generation of satellites, but partnership between space-based observation and accurate ground-based in situ measurement promises much better understanding of global carbon budgets.

Measurements of greenhouse gases in the air and knowledge of the winds and atmospheric transport are used in atmospheric transport inversions to test and assess emissions. Modeling studies based on records from many stations can quantify emissions over large regions, from nation-scale to continents to global (5, 9, 13, 14). In western Europe, with the existing network of measurement stations, it may be possible to recover CO_2 fluxes reliably on a scale of 1000 km and over a time period of 10 days (15, 16). This is close to what is needed for validating bottom-up emissions

(which are quoted on an annual basis). In the United States, NOAA's CarbonTracker (17) is a community tool that aims to produce quantitative estimates of atmospheric carbon uptake and release for North America and the rest of the world that are consistent with observed patterns of CO_2 in the atmosphere.

New modeling methods are also being developed that will, for example, allow easy nesting of high-resolution regional models into global-scale models of atmospheric transport (18). For CO_2 , much needs to be done to quantify biological fluxes (19), but this type of modeling is already being applied to CH_4 and some industrial greenhouse gases. In Europe, large-scale inversion modeling of CH_4 has been used to challenge some national emissions inventories (20).

The pressures of emissions control legislation and the value of emissions reductions in carbon-equivalent trading markets create strong incentives for intentional underreporting and for inequity among nations. Legislators need to acknowledge that large uncertainties in emissions exist, and promote advances in both bottom-up and top-down assessments so that they agree within specified uncertainties. Perhaps 5% would be a realistic initial target.

Cooperative global monitoring is not new. The Comprehensive Test Ban Treaty Organization has undertaken a program to locate and quantify atmospheric emissions of radioisotopes based on a network of ground-based measurement stations and inverse modeling of atmospheric transport. Why not begin by adding greenhouse gas measurements to these stations? The successful use of this approach to monitor greenhouse gas emissions at national and subnational scales will surely demand a much better global network of high-frequency in situ measurements than we currently have, but the investment would not be great compared to the economic cost of failed regulation.

Indeed, compared to the scale of the climate problem, in situ measurements, and even satellites, are relatively small investments. But verification of emissions demands a sustained multiyear effort that can be anathema to current approaches to research support, especially as the work has a strong discovery component. It also needs to be supported, at least in part, through peer-reviewed research channels. Rigorous quality control is essential if data from different methods and programs are to be adequately integrated. The results will be accurate within modeled error, based on direct

measurement, and independent of reporting bias. Such verification may be critical to a future comprehensive convention on climate change.

References

1. Committee on Methods for Estimating Greenhouse Gas Emissions, *Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements* (National Research Council, Washington, DC, 2010).
2. Royal Astronomical Society Workshop Report, *Verification of Greenhouse Gas Emissions by Atmospheric Measurement*; www.ras.org.uk/index.php?option=com_content&task=view&id=936&Itemid=12.
3. M. Marquis, P. Tans, *Science* **320**, 460 (2008).
4. I. Levin *et al.*, *Atmos. Chem. Phys.* **10**, 2655 (2010).
5. A. Stohl *et al.*, *Atmos. Chem. Phys.* **9**, 1597 (2009).
6. J. Mühle *et al.*, *Atmos. Chem. Phys. Discuss.* **10**, 6485 (2010).
7. R. F. Weiss, J. Mühle, P. K. Salameh, C. M. Harth, *Geophys. Res. Lett.* **35**, L20821 (2008).
8. C. D. Keeling, *Tellus* **12**, 200 (1960).
9. I. Pison, P. Bousquet, F. Chevallier, S. Szopa, D. Hauglustaine, *Atmos. Chem. Phys.* **9**, 5281 (2009).
10. D. Lowry, C. W. Holmes, N. D. Rata, P. O'Brien, E. G. Nisbet, *J. Geophys. Res.* **106** (D7), 7427 (2001).
11. I. Levin, C. Rödenbeck, *Naturwissenschaften* **95**, 203 (2008).
12. B. B. Stephens *et al.*, *Science* **316**, 1732 (2007).
13. P. P. Tans *et al.*, *Science* **247**, 1431 (1990).
14. Y.-H. Chen, R. G. Prinn, *J. Geophys. Res.* **111** (D10), D10307 (2006).
15. C. Carouge, P. Bousquet, P. Peylin, P. Rayner, P. Ciais, *Atmos. Chem. Phys.* **10**, 3107 (2010).
16. C. Carouge *et al.*, *Atmos. Chem. Phys.* **10**, 3119 (2010).
17. National Oceanic and Atmospheric Administration, www.esrl.noaa.gov/gmd/ccgg/carbontracker/index.html.
18. C. Rödenbeck, C. Gerbig, K. Trusilova, M. Heimann, *Atmos. Chem. Phys.* **9**, 5331 (2009).
19. M. Vetter *et al.*, *Biogeosciences* **5**, 561 (2008).
20. P. Bergamaschi *et al.*, *Atmos. Chem. Phys.* **5**, 2431 (2005).

10.1126/science.1189936

VIROLOGY

Helping the Resistance

Edward C. Holmes^{1,2}

The evolution of antiviral drug resistance sounds like a simple Darwinian story. The high mutation rate that characterizes RNA viruses ensures that drug-resistant mutations are generated continuously, and the global use of antivirals provides the selection pressure for these mutations to sweep through viral populations. In some cases, however, reality is more complex. The mutations that confer antiviral resistance may have a detrimental effect on viral fitness in the absence of the drug so that secondary fitness-restoring mutations must occur to enable the large-scale spread of resistance. More puzzling, drug resistance can also occur in the absence of the main agent of selection—the widespread use of antiviral drugs. Both of these evolutionary conundrums are apparent in one of the most important cases of antiviral resistance in recent years—the global spread of resistance to oseltamivir in seasonal H1N1 influenza A virus. On page 1272 of this issue, Bloom *et al.* (1) show that in a “permissive” genetic background, seasonal H1N1 virus avoids the fitness costs normally associated with oseltamivir resistance.

The evolutionary ancestry of seasonal H1N1 influenza A virus lies with the devastating pandemic of 1918. Although the

1918 virus eventually died out in the 1950s, the H1N1 lineage reappeared in 1977 and has since circulated with H3N2 influenza A virus, and more recently with the newly emerged swine-origin pandemic H1N1/09 virus. Although seasonal H1N1 is associated with lower mortality than either the H3N2 or pandemic H1N1/09 viruses, in some years it is the dominant strain infecting humans and a considerable public health concern. It was therefore worrying that in the winter of 2007–2008, the influenza season in the Northern Hemisphere, reports of oseltamivir resistance in seasonal H1N1 began to appear in northern Europe and in patients who had not received

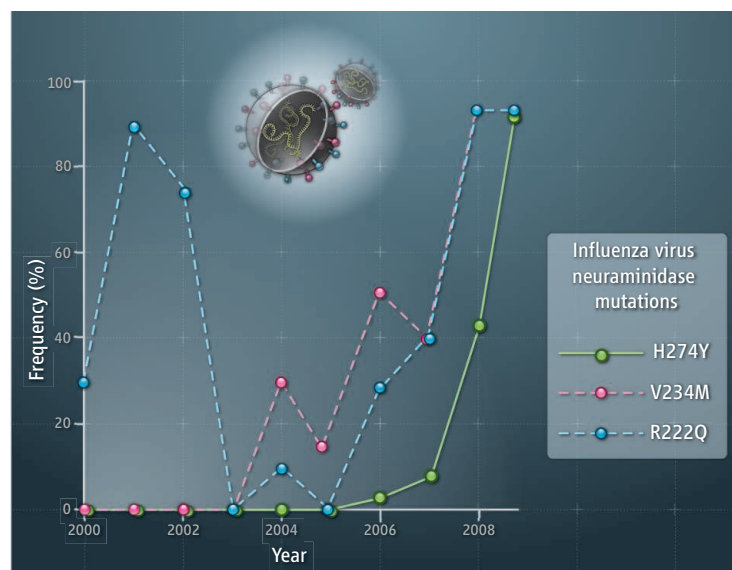
Phylogenetic analysis reveals mutations that led to the rise of drug resistance in seasonal H1N1 influenza.

oseltamivir (2, 3). By the following 2008–2009 influenza season, it was clear that seasonal H1N1 virus was resistant to oseltamivir on a global scale (4–6), effectively ruling out further use of this frontline antiviral drug.

Viral neuraminidase cleaves terminal sialic acid residues from glycoproteins on the surface of an infected host cell, thereby promoting the release of progeny viruses. Oseltamivir binds to the neuraminidase and blocks this function. Resistance to oseltamivir is often associated with a histidine (H) to tyrosine (Y) amino acid change at residue 274 (H274Y) of the influenza neuraminidase. Bloom *et al.* reaffirmed that in the absence

of oseltamivir the H274Y mutation has a detrimental effect on viral fitness. For the mutation to become “fixed” on a global scale—that is, occur at 100% frequency in the viral population—it is therefore essential that the negative fitness effect is offset by secondary mutations located elsewhere in the viral genome. Bloom *et al.* used a phylogenetic analysis of H1N1 gene sequences to identify those amino acid changes that occurred in the viral lineage that eventually developed oseltamivir resistance, and hence before the H274Y mutation appeared.

This analysis revealed two amino acid substitutions in the neuraminidase protein that seemed to pre-empt H274Y—replacement



The rise of oseltamivir resistance. The graph depicts the changing global frequency of neuraminidase mutations associated with oseltamivir resistance in seasonal H1N1 influenza A virus from 2000 to 2009. The H274Y mutation is the main resistant determinant. Mutations V234M and R222Q arose earlier and were permissive to the spread of H274Y. Numbers were calculated from all H1N1 neuraminidase sequences available at GenBank.

¹Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA. ²Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA. E-mail: ech15@psu.edu

of valine (V) with methionine (M) at residue 234 (V234M) and replacement of arginine (R) with glutamine (Q) at residue 222 (R222Q) (see the figure). The authors confirmed through a series of fitness experiments that V234M and R222Q did indeed restore the fitness of H274Y-bearing viruses to that observed in wild-type viruses.

This powerful synthesis of evolutionary and laboratory investigation has a number of important implications. Whereas secondary fitness-restoring changes are normally considered as occurring after the mutation of interest, Bloom *et al.* show that secondary mutations can also precede the major change of phenotype. This, in turn, means that some viral lineages are inherently better able to evolve important phenotypic traits than others. Although Bloom *et al.* focused on the evolution of oseltamivir resistance, in theory the enabling effect of permissive mutations could extend to the appearance of lineages with altered antigenicity, or to those that allow viruses to adapt to new host species. It is also possible that the existence of permissive mutations provides some power to predict which viral lineages are likely to evolve drug resistance in the future. If we know that a particular viral lineage harbors the V234M and R222Q mutations, it would seem advisable to spend more time surveying these lineages for future drug resistance

than those in which the permissive mutations are absent (although it is clearly the case that other mutations could be permissive in different genetic backgrounds). The most important arena for such evolutionary forecasting involves the pandemic H1N1/09 virus, which is still the most common influenza A virus circulating today. The good news is that to date H274Y has made only a sporadic appearance in these viruses; of the more than 2500 pandemic H1N1/09 neuraminidase sequences available in the sequence database GenBank, only about 1% carry H274Y. In addition, neither V234M nor R222Q has yet been detected in pandemic H1N1/09 viruses, the vast majority of which possess an asparagine (N) amino acid at site 222. It is possible that H274Y has not yet spread globally in pandemic H1N1/09 because it lacks the essential permissive mutations. The appearance of V234M and R222Q in pandemic H1N1/09 viruses may therefore herald the onset of oseltamivir resistance.

A looming question is why H274Y started to spread at all. Although it is an effective antiviral agent, there does not appear to have been sufficient use of oseltamivir to provide a selection pressure strong enough for the global fixation of H274Y (6, 7). One possible explanation is that H274Y became fortuitously linked to a major fitness-enhancing mutation located elsewhere in the viral

genome, such that it was swept to fixation by a process of “genetic hitchhiking” rather than by direct selection. Indeed, such fortuitous linkage was likely responsible for the dramatic rise of resistance to the antiviral drug adamantane in seasonal H3N2 viruses (8). However, an analysis of the complete genomes of seasonal H1N1 viruses available to date provides no evidence for hitchhiking. Because neuraminidase is involved in a number of fundamental virological processes, including immune escape, it is possible that the V234M and R222Q mutations not only restore viral fitness in the presence of H274Y, but also extend it beyond what is seen in wild-type viruses. Assessing the fitness of V234M, R222Q, and H274Y mutations in animal models may provide the necessary test of this hypothesis. Until this happens, why oseltamivir resistance spread globally with such speed will remain an intriguing evolutionary puzzle.

References

1. J. D. Bloom *et al.*, *Science* **328**, 1272 (2010).
2. A. Lackenby *et al.*, *Euro Surveill.* **13**, 8026 (2008).
3. A. Nicoll *et al.*, *Euro Surveill.* **13**, 8025 (2008).
4. B. Ciancio *et al.*, *Euro Surveill.* **14**, 19412 (2009).
5. T. Baranovich *et al.*, *J. Clin. Virol.* **47**, 23 (2010).
6. N. J. Dharan *et al.*, *JAMA* **301**, 1034 (2009).
7. P. Kramarz, D. Monnet, A. Nicoll, C. Yilmaz, B. Ciancio, *Euro Surveill.* **14**, 19112 (2009).
8. L. Simonsen *et al.*, *Mol. Biol. Evol.* **24**, 1811 (2007).

10.1126/science.1190994

CELL SIGNALING

Anchors Away for Ubiquitin Chains

Kislay Parvatiyar and Edward W. Harhaj

It has become increasingly clear that cellular proteins can become conjugated with chains of the small protein ubiquitin without becoming degraded by the proteasome. This “noncanonical” process allows ubiquitinated proteins to regulate numerous processes including DNA repair and cell signaling. However, determining the mechanisms by which polyubiquitin chains affect such processes has been challenging, in part because of the unavailability of genetically engineered mice that lack ubiquitin and the partial redundancies of molecules that support ubiquitin chain assembly. Recent work now shows that rather than being assembled onto target proteins by the cell’s ubiquitin con-

jugation machinery, ubiquitin is also present in cells as free, unanchored chains that can bind directly to a signaling protein involved in the immune response to viral infection (1). This raises the question of whether unanchored ubiquitin chains generally regulate other physiological functions.

The retinoic acid-inducible gene-I (RIG-I)-like family of cytosolic pattern recognition receptors play an essential role in initiating the host antiviral response by detecting nucleic acids derived from viral genomes. Upon binding to viral RNAs that bear a 5'-triphosphate, RIG-I signals to the transcription factor interferon regulatory factor 3 (IRF3), which upon activation elicits an antiviral program to restrict virus replication and spread. RIG-I signaling to IRF3 is mediated by mitochondria antiviral signaling protein (MAVS; also known as IPS-1, VISA,

Unanchored chains of ubiquitin molecules control a cell signaling pathway that responds to viral infection.

and Cardif). MAVS further recruits adaptor proteins and enzymes (TBK1 and IKK ϵ) to phosphorylate IRF3, ultimately resulting in dimerization of the transcription factor and subsequent nuclear translocation. Zeng *et al.* (2) previously showed that noncanonical protein ubiquitination is essential for activating IRF3 via MAVS, but the connection between these events was not clear. Using an in vitro cell-free system, Zeng *et al.* (1) recapitulated RIG-I-IRF3 signaling, showing that virally activated RIG-I, together with a mitochondrial fraction (containing MAVS) and cytosolic extracts from cells (containing TBK1 and IKK ϵ kinases), support IRF3 activation. They further demonstrated in this system that the binding of a viral RNA mimic to purified RIG-I was not sufficient to activate IRF3, but also required ubiquitination of RIG-I. The latter event required the E3 ubiquitin

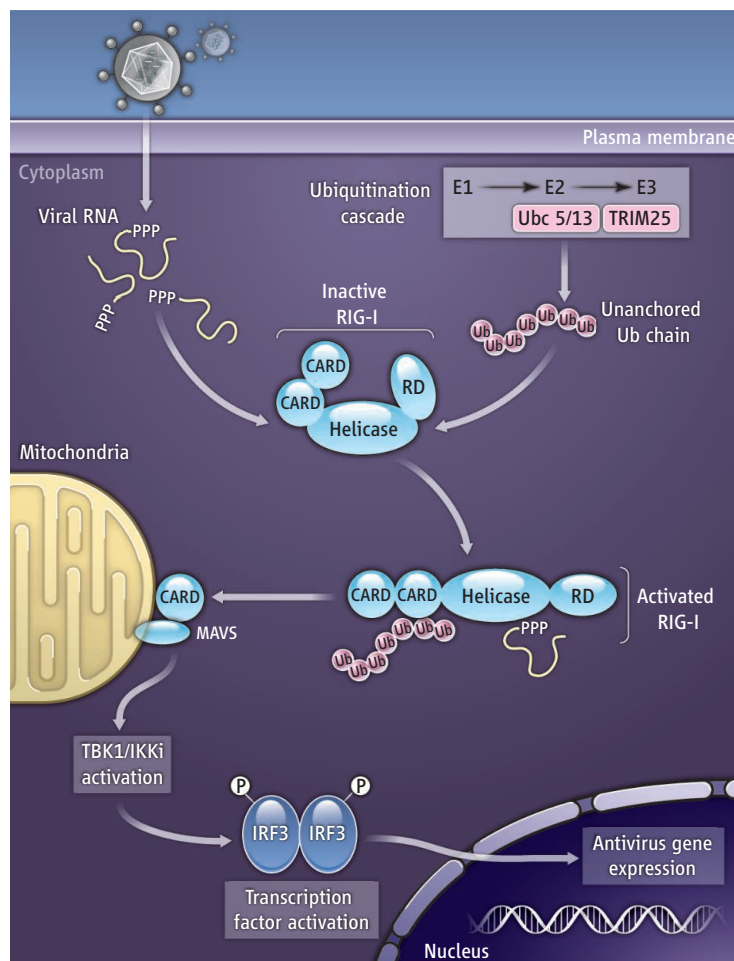
Department of Microbiology and Immunology, Sylvester Comprehensive Cancer Center, Miller School of Medicine, University of Miami, Miami, FL 33136, USA. E-mail: eharhaj@med.miami.edu

ligase TRIM25, an enzyme that functions in the cell's ubiquitin conjugation (Ubc) mechanism. TRIM25 was previously shown to potentiate RIG-I signaling by promoting its ubiquitination (3).

Two caspase recruitment domains (CARDs) at the amino terminus of RIG-I [RIG-I(N)] are necessary and sufficient to activate IRF3 independently of viral RNA (4). Because these domains are also the target of polyubiquitin chain addition (catalyzed by TRIM25) to stimulate RIG-I signaling, Zeng *et al.* examined constitutively active RIG-I(N) to tease out how RIG-I activation signals to MAVS in a manner that depends on ubiquitin. In agreement with previous studies (3), RIG-I(N) required ubiquitination to activate IRF3 in the in vitro cell-free system. The ubiquitin moiety consists of 76 amino acids, including 7 lysine residues. Chains of ubiquitin may arise through linkage at the lysine residue in position 63 of the polypeptide (Lys⁶³). When the authors incubated purified RIG-I(N) with preassembled Lys⁶³-linked polyubiquitin chains, activation of IRF3 was achieved, which suggested that RIG-I(N) did not require assembly of a ubiquitin chain onto it by the cellular conjugation machinery, but direct ubiquitin chain binding instead to activate IRF3. Thus, the CARDs of RIG-I serve as ubiquitin sensors that recognize unanchored Lys⁶³-linked polyubiquitin chains synthesized by TRIM25.

Zeng *et al.* (1) also found that viral RNA binding was a required prerequisite for RIG-I binding to unanchored polyubiquitin chains, supporting a two-step model in which viral RNA binding to the regulatory domain of RIG-I promotes a conformational change that allows unanchored polyubiquitin chains to bind to the CARDs (see the figure). Conceivably, RIG-I binding to ubiquitin may also control interactions with other effectors such as stimulator of interferon genes (STING), an endoplasmic reticulum protein that binds to both RIG-I and MAVS (5).

The powerful nature of the cell-free system employed by Zeng *et al.* (1) establishes a potent tool for analyzing cellular signaling. However, the authors noted some curious findings. Although the polyubiquitin chains



Unanchored in defense. During virus infection, viral RNA is detected by the regulatory domain (RD) of RIG-I, triggering a conformational change that allows an open conformation. The CARDs of RIG-I bind to unanchored Lys⁶³-linked polyubiquitin (Ub) chains. This facilitates RIG-I interaction with MAVS protein in the mitochondria and subsequent signaling to IRF3, which controls the expression of antiviral genes.

can assemble onto RIG-I through the ubiquitin conjugation machinery, the reason for modification through this route is unclear, as it is dispensable for IRF3 activation. One possibility is that the polyubiquitin chains anchored onto RIG-I through the conjugation system somehow mark RIG-I for negative feedback regulation by recruiting deubiquitinases for signal termination (6).

Zeng *et al.* (1) also noted that the CARDs of either MAVS or the intracellular pattern recognition receptor NOD2 do not bind ubiquitin. Because NOD2 also uses MAVS to activate IRF3 in response to viral RNA (7), it is of interest to determine why ubiquitin regulates the RIG-I-MAVS signaling pathway but not the NOD2-MAVS pathway in response to viral infection. The authors, however, did observe that the CARDs of the RIG-I homolog MDA5 bind directly to polyubiquitin chains, which suggests that MDA5 and RIG-I may be regulated in a similar manner (1). A next step is to identify whether

TRIM25 or another ubiquitin ligase synthesizes the unanchored polyubiquitin chains for MDA5.

TRIM25 has also been implicated in modifying proteins with the ubiquitin-like protein ISG15 (8). Conjugation of ISG15 onto RIG-I inhibits its signaling through a negative feedback loop (9). Thus, a more refined analysis is needed to understand the complex role of TRIM25 in antiviral signaling, perhaps achieved with a modified in vitro system containing both ubiquitination and ISGylation conjugation systems.

Do unanchored polyubiquitin chains play a pervasive role in regulating many signaling pathways? Indeed, Xia *et al.* recently reported that unanchored Lys⁶³-linked polyubiquitin chains play an important role in activating the TAK1 and IKK kinases in the NF- κ B signaling pathway (10). Zeng *et al.* (1) estimate approximately 6000 molecules of unanchored Lys⁶³-linked polyubiquitin chains per cell (the authors studied a human embryonic kidney cell line), although RIG-I may sense as few as 15 molecules of unanchored

polyubiquitin chains. It is likely that the number of unanchored polyubiquitin chains in a given cell is determined by the balance of ubiquitin ligases and deubiquitinases, which in turn is regulated by specific signaling pathways. Unanchored polyubiquitin chains may also be regulated by localization and compartmentalization within the cell, adding yet another dimension of complexity.

References and Notes

1. W. Zeng *et al.*, *Cell* **141**, 315 (2010).
2. W. Zeng, M. Xu, S. Liu, L. Sun, Z. J. Chen, *Mol. Cell* **36**, 315 (2009).
3. M. U. Gack *et al.*, *Nature* **446**, 916 (2007).
4. M. Yoneyama *et al.*, *Nat. Immunol.* **5**, 730 (2004).
5. H. Ishikawa, G. N. Barber, *Nature* **455**, 674 (2008).
6. M. Zhang *et al.*, *J. Biol. Chem.* **283**, 18621 (2008).
7. A. Sabbah *et al.*, *Nat. Immunol.* **10**, 1073 (2009).
8. W. Zou, D. E. Zhang, *J. Biol. Chem.* **281**, 3989 (2006).
9. M. J. Kim, S. Y. Hwang, T. Imaizumi, J. Y. Yoo, *J. Virol.* **82**, 1474 (2008).
10. Z. P. Xia *et al.*, *Nature* **461**, 114 (2009).
11. E.W.H. is funded by NIH grants R01CA135362, R01GM083143, and P01CA128115.

10.1126/science.1192296

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
<http://www.sciencemag.org/cgi/content/full/328/5983/1246>

Supporting Online Material can be found at:
<http://www.sciencemag.org/cgi/content/full/328/5983/1246/DC1>

This article **cites 28 articles**, 2 of which can be accessed for free:
<http://www.sciencemag.org/cgi/content/full/328/5983/1246#otherarticles>

This article appears in the following **subject collections**:
Physics
<http://www.sciencemag.org/cgi/collection/physics>

Highly Mobile Gapless Excitations in a Two-Dimensional Candidate Quantum Spin Liquid

Minoru Yamashita,^{1*} Norihito Nakata,¹ Yoshinori Senshu,¹ Masaki Nagata,¹ Hiroshi M. Yamamoto,^{2,3} Reizo Kato,² Takasada Shibauchi,¹ Yuji Matsuda^{1*}

The nature of quantum spin liquids, a novel state of matter where strong quantum fluctuations destroy the long-range magnetic order even at zero temperature, is a long-standing issue in physics. We measured the low-temperature thermal conductivity of the recently discovered quantum spin liquid candidate, the organic insulator $\text{EtMe}_3\text{Sb}[\text{Pd}(\text{dmit})_2]_2$. A sizable linear temperature dependence term is clearly resolved in the zero-temperature limit, indicating the presence of gapless excitations with an extremely long mean free path, analogous to excitations near the Fermi surface in pure metals. Its magnetic field dependence suggests a concomitant appearance of spin-gap-like excitations at low temperatures. These findings expose a highly unusual dichotomy that characterizes the low-energy physics of this quantum system.

Spin systems confined to low dimensions exhibit a rich variety of quantum phenomena. Particularly intriguing are quantum spin liquids (QSLs), antiferromagnets with quantum fluctuation-driven disordered ground states, which have been attracting tremendous attention for decades (1). The notion of QSLs is now firmly established in one-dimensional (1D) spin systems. In dimensions greater than one, it is widely believed that QSL ground states emerge when interactions among the magnetic degrees of freedom are incompatible with the underlying crystal geometry, leading to a strong enhancement of quantum fluctuations. In 2D, typical examples of systems where such geometrical frustrations are present are the triangular and kagomé lattices. Largely triggered by the proposal of the resonating-valence-bond theory on a 2D triangular lattice and its possible application to high-transition temperature cuprates (2), realizing QSLs in 2D systems has been a long-sought goal. However, QSL states are hard to achieve experimentally because the presence of small but finite 3D magnetic interactions usually results in some ordered (or frozen) state. Two recently discovered organic insulators, κ -[bis(ethylenedithio)-tetrathiafulvalene]₂Cu₂(CN)₃ [κ -(BEDT-TTF)₂Cu₂(CN)₃] (3) and $\text{EtMe}_3\text{Sb}[\text{Pd}(\text{dmit})_2]_2$ (4, 5), both featuring 2D spin-1/2 Heisenberg triangular lattices, are believed to be promising candidate materials that are likely to host QSLs. In both compounds, nuclear magnetic resonance (NMR) measurements have shown no

long-range magnetic order down to a temperature corresponding to $J/12,000$, where J (~250 K for both compounds) is the nearest-neighbor spin interaction energy (exchange coupling) (3, 5). In a triangular lattice antiferromagnet, the frustration brought on by the nearest-neighbor Heisenberg

interaction is known to be insufficient to destroy the long-range ordered ground state (6). This has led to the proposals of numerous scenarios which might stabilize a QSL state: spinon Fermi surface (7, 8), algebraic spin liquid (9), spin Bose metal (10), ring-exchange model (11), Z_2 spin liquid state (12), chiral spin liquid (13), Hubbard model with a moderate onsite repulsion (14, 15), and one-dimensionalization (16, 17). Nevertheless, the origin of the QSL in the organic compounds remains an open question.

To understand the nature of QSLs, knowledge of the detailed structure of the low-lying elementary excitations in the zero-temperature limit, particularly the presence or absence of an excitation gap, is of primary importance (18). Such information bears immediate implications on the spin correlations of the ground state, as well as the correlation length scale of the QSL. For example, in 1D spin-1/2 Heisenberg chains, the elementary excitations are gapless spinons (chargeless spin-1/2 quasiparticles) characterized by a linear energy dispersion and a power-law decay of the spin correlation (19), whereas in the integer spin case such excitations are gapped (20). In the organic compound κ -(BEDT-TTF)₂Cu₂(CN)₃, where the first putative QSL state was reported (3), the presence of the spin excitation gap is controversial (18, 21). In this compound, the stretched, non-

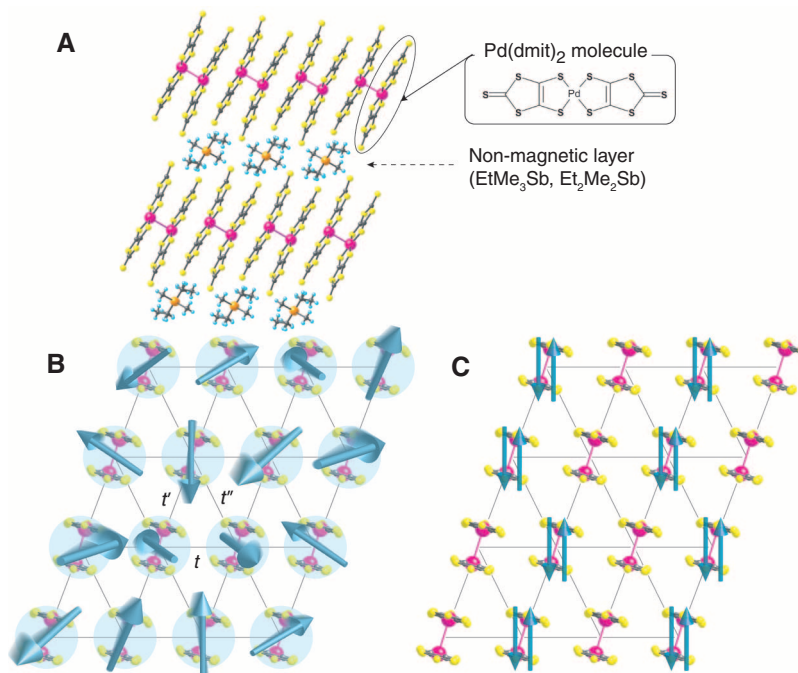


Fig. 1. The crystal structure of $\text{EtMe}_3\text{Sb}[\text{Pd}(\text{dmit})_2]_2$ and $\text{Et}_2\text{Me}_2\text{Sb}[\text{Pd}(\text{dmit})_2]_2$. (A) A view parallel to the 2D magnetic $\text{Pd}(\text{dmit})_2$ layer, separated by layers of a nonmagnetic cation. (B) The spin structure of the 2D planes of $\text{EtMe}_3\text{Sb}[\text{Pd}(\text{dmit})_2]_2$ (dmit-131), where $\text{Et} = \text{C}_2\text{H}_5$, $\text{Me} = \text{CH}_3$, and $\text{dmit} = 1,3$ -dithiole-2-thione-4,5-dithiolate. $\text{Pd}(\text{dmit})_2$ are strongly dimerized (table S1), forming spin-1/2 units $[\text{Pd}(\text{dmit})_2]_2^-$ (blue arrows). The antiferromagnetic frustration gives rise to a state in which none of the spins are frozen down to 19.4 mK (4). (C) The spin structure of the 2D planes of $\text{Et}_2\text{Me}_2\text{Sb}[\text{Pd}(\text{dmit})_2]_2$ (dmit-221). A charge order transition occurs at 70 K, and the units are separated as neutral $[\text{Pd}(\text{dmit})_2]_2^0$ and divalent dimers $[\text{Pd}(\text{dmit})_2]_2^{2-}$. The divalent dimers form intradimer valence bonds, showing a nonmagnetic spin singlet (blue arrows) ground state with a very large excitation gap (24).

¹Department of Physics, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan. ²RIKEN, Wako-shi, Saitama 351-0198, Japan. ³Japan Science and Technology Agency, Precursory Research for Embryonic Science and Technology (JST-PRESTO), Kawaguchi, Saitama 332-0012, Japan.

*To whom correspondence should be addressed. E-mail: yamashitamino@scphys.kyoto-u.ac.jp (M.Y.); matsuda@scphys.kyoto-u.ac.jp (Y.M.)

exponential decay of the NMR relaxation indicates inhomogeneous distributions of spin excitations (22), which may obscure the intrinsic properties of the QSL. A phase transition possibly associated with the charge degree of freedom at ~ 6 K further complicates the situation (23). Meanwhile, in $\text{EtMe}_3\text{Sb}[\text{Pd}(\text{dmit})_2]_2$ (dmit-131) such a transition is likely to be absent, and a much more homogeneous QSL state is attained at low temperatures (4, 5). As a further merit, dmit-131 (Fig. 1B) has a cousin material $\text{Et}_2\text{Me}_2\text{Sb}[\text{Pd}(\text{dmit})_2]_2$ (dmit-221) with a similar crystal structure (Fig. 1C), which exhibits a nonmagnetic charge-ordered state with a large excitation gap below 70 K (24). A comparison between these two related materials will therefore offer us the opportunity to single out genuine features of the QSL state believed to be realized in dmit-131.

Measuring thermal transport is highly advantageous for probing the low-lying elementary excitations in QSLs, because it is free from the nuclear Schottky contribution that plagues the heat capacity measurements at low temperatures (21). Moreover, it is sensitive exclusively to itinerant spin excitations that carry entropy, which provides important information on the nature of the

spin correlation and spin-mediated heat transport. Indeed, highly unusual transport properties including the ballistic energy propagation have been reported in a 1D spin-1/2 Heisenberg system (25).

The temperature dependence of the thermal conductivity κ_{xx} divided by T of a dmit-131 single crystal displays a steep increase followed by a rapid decrease after showing a pronounced maximum at $T_g \sim 1$ K (Fig. 2A). The heat is carried primarily by phonons (κ_{xx}^{ph}) and spin-mediated contributions ($\kappa_{xx}^{\text{spin}}$). The phonon contribution can be estimated from the data of the nonmagnetic state in a dmit-221 crystal with similar dimensions, which should have a negligibly small $\kappa_{xx}^{\text{spin}}$. In dmit-221, $\kappa_{xx}^{\text{ph}}/T$ exhibits a broad peak at around 1 K, which appears when the phonon conduction grows rapidly and is limited by the sample boundaries. On the other hand, κ_{xx}/T of dmit-131, which well exceeds $\kappa_{xx}^{\text{ph}}/T$ of dmit-221, indicates a substantial contribution of spin-mediated heat conduction below 10 K. This observation is reinforced by the large magnetic field dependence of κ_{xx} of dmit-131, as discussed below (Fig. 3A). Figure 2B shows a peak in the κ_{xx} versus T plot for dmit-131, which is absent in dmit-221. We therefore conclude that $\kappa_{xx}^{\text{spin}}$ and $\kappa_{xx}^{\text{spin}}/T$ in dmit-131 have

a peak structure at $T_g \sim 1$ K, which characterizes the excitation spectrum.

The low-energy excitation spectrum can be inferred from the thermal conductivity in the low-temperature regime. In dmit-131, κ_{xx}/T at low temperatures is well fitted by $\kappa_{xx}/T = \kappa_{00}/T + bT^2$ (Fig. 2C), where b is a constant. The presence of a residual value in κ_{xx}/T at $T \rightarrow 0$ K, κ_{00}/T , is clearly resolved. The distinct presence of a nonzero κ_{00}/T term is also confirmed by plotting κ_{xx}/T versus T (Fig. 2D). In sharp contrast, in dmit-221, a corresponding residual κ_{00}/T is absent and only a phonon contribution is observed (26). The residual thermal conductivity in the zero-temperature limit immediately implies that the excitation from the ground state is gapless, and the associated correlation function has a long-range algebraic (power-law) dependence. We note that the temperature dependence of κ_{xx}/T in dmit-131 is markedly different from that in $\kappa\text{-(BEDT-TTF)}_2\text{Cu}_2(\text{CN})_3$, in which the exponential behavior of κ_{xx}/T associated with the formation of excitation gap is observed (18).

Key information on the nature of elementary excitations is further provided by the field dependence of κ_{xx} . Because it is expected that κ_{xx}^{ph} is hardly influenced by the magnetic field, particularly at very low temperatures, the field dependence is governed by $\kappa_{xx}^{\text{spin}}(H)$ (26). The obtained H -dependence, $\kappa_{xx}(H)$, at low temperatures is quite unusual (Fig. 3A). At the lowest temperature, $\kappa_{xx}(H)$ at low fields is insensitive to H but displays a steep increase above a characteristic magnetic field $H_g \sim 2$ T. At higher temperatures close to T_g , this behavior is less pronounced, and at 1 K $\kappa_{xx}(H)$ increases with H nearly linearly. The observed field dependence implies that some spin-gap-like excitations are also present at low temperatures, along with the gapless excitations inferred from the residual κ_{00}/T . The energy scale of the gap is characterized by $\mu_B H_g$, which is comparable to $k_B T_g$. Thus, it is natural to associate the observed zero-field peak in $\kappa_{xx}(T)/T$ at T_g with the excitation gap formation.

Next we examined a dynamical aspect of the spin-mediated heat transport. An important question is whether the observed energy transfer via elementary excitations is diffusive or ballistic. In the 1D spin-1/2 Heisenberg system, the ballistic energy propagation occurs as a result of the conservation of energy current (25). Assuming the kinetic approximation, the thermal conductivity is written as $\kappa_{xx}^{\text{spin}} = C_s v_s \ell_s / 3$, where C_s is the specific heat, v_s is the velocity, and ℓ_s is the mean free path of the quasiparticles responsible for the elementary excitations. We tried to estimate ℓ_s simply by assuming that the linear term in the thermal conductivity arises from the fermionic excitations, in analogy with excitations near the Fermi surface in metals. The residual term is written as $\kappa_{00}/T \sim (k_B^2 / d a \hbar) \ell_s$, where d (~ 3 nm) and a (~ 1 nm) are interlayer and nearest-neighbor spin distance. We assumed the linear energy dispersion $\epsilon(k) = \hbar v_s k$, a 2D density of states and a Fermi energy comparable to J (26). From the observed κ_{00}/T , we find that ℓ_s reaches as long as ~ 1 μm , indicating

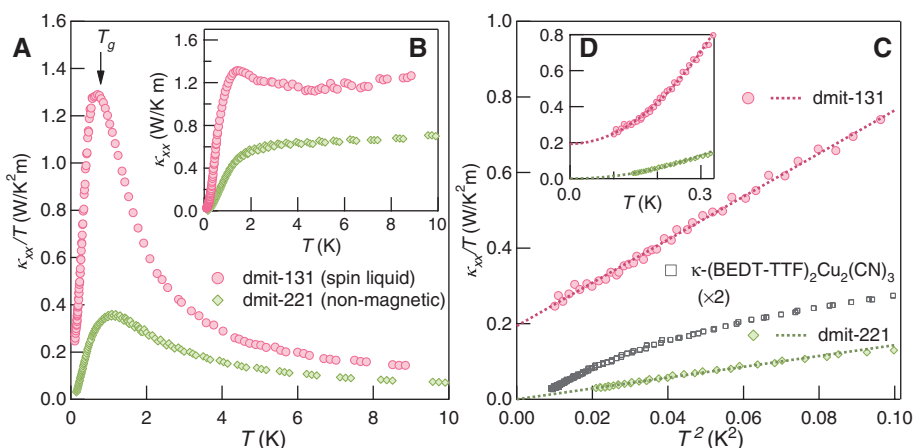
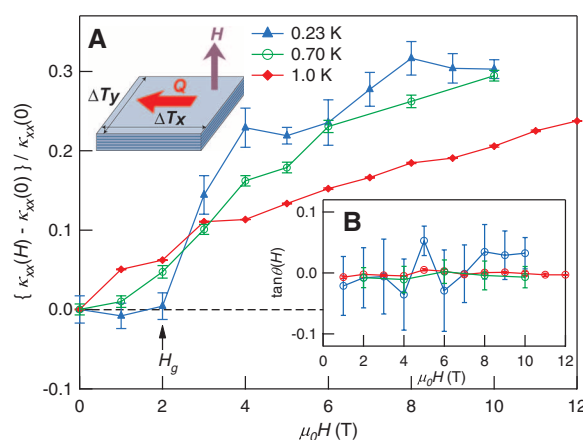


Fig. 2. The temperature dependence of $\kappa_{xx}(T)/T$ (A) and $\kappa_{xx}(T)$ (B) of dmit-131 (pink) and dmit-221 (green) below 10 K in zero field [$\kappa_{xx}(T)$ is the thermal conductivity]. A clear peak in κ_{xx}/T is observed in dmit-131 at $T_g \sim 1$ K, which is also seen as a hump in κ_{xx} . Lower temperature plot of $\kappa_{xx}(T)/T$ as a function of T^2 (C) and T (D) of dmit-131, dmit-221, and $\kappa\text{-(BEDT-TTF)}_2\text{Cu}_2(\text{CN})_3$ (black) (18). A clear residual of $\kappa_{xx}(T)/T$ is resolved in dmit-131 in the zero-temperature limit.

Fig. 3. (A) Field dependence of thermal conductivity normalized by the zero field value, $[\kappa_{xx}(H) - \kappa_{xx}(0)]/\kappa_{xx}(0)$ of dmit-131 at low temperatures. (Inset) The heat current $\mathbf{Q}$ was applied within the 2D plane, and the magnetic field $\mathbf{H}$ was perpendicular to the plane. κ_{xx} and κ_{xy} were determined by diagonal and off-diagonal temperature gradients, ΔT_x and ΔT_y , respectively. (B) Thermal-Hall angle $\tan\theta(H) = \kappa_{xy}/(\kappa_{xx} - \kappa_{xx}^{\text{ph}})$ as a function of H at 0.23 K (blue), 0.70 K (green), and 1.0 K (red).



that the excitations are mobile to a distance 1000 times as long as the interspin distance without being scattered. Remarkably, the observed ℓ_s is even longer than the mean impurity distance estimated from the susceptibility measurements (5). Although our simple estimation should be scrutinized, the nearly ballistic propagation seems to be realized in this QSL state, which bears a striking resemblance to the 1D Heisenberg system (25). Such a coherent motion of the excitation is only possible with an extremely long correlation length, consistent with the power-law correlation function with gapless excitations.

Our results indicate that there are two kinds of excitations in the low-temperature regime of dmit-131: One is a low-lying gapless excitation and another is a spin-gap-like excitation that couples to magnetic fields. The existence of the gapless mode imposes constraints on theories that can account for the QSL in this material. For instance, it contradicts the identification of the ground state of this system with a fully gapped short-range resonating-valence-bond state (27). It is intriguing that, although there has been no experimental observation, the coexistence of the gapless and gapped excitations has been theoretically predicted in highly frustrated kagomé lattice (28). In this case, the magnetic excitations are separated from the ground state by a spin-gap, which is filled with nonmagnetic excitations. It is a nontrivial problem, however, if such a state can be realized in a triangular lattice where less quantum degenerate states are expected.

Recently, the ground state of a QSL in the 2D triangular lattice has been discussed in terms of the spinon Fermi surface (7, 8). The Fermi surface is suggested to have pairing instability at low temperatures, and a possible nodal gap structure appears that is analogous to d -wave superconductivity in high-transition temperature cuprates (29). In this scenario, the spin-gap-like behavior is at-

tributed to the pairing gap formation, and the finite residual T -linear term stems from the zero-energy density of states similar to the disorder-induced normal fluid in d -wave superconductors. The low energy density of states is expected to be insensitive to Zeeman magnetic field as in a Fermi liquid. The nodal gap is expected to produce thermally excited quasiparticles, which may account for the observed T^2 dependence of κ/T with a much steeper slope than κ_{xx}^{ph}/T (Fig. 2C). This model also predicts a thermal-Hall effect of spinons that experience the Lorentz force, akin to the conduction electrons in metals (30). In an attempt to observe this effect, we measured the thermal-Hall conductivity κ_{xy} and the tangent of the thermal-Hall angle, $\tan\theta(H) = \kappa_{xy}/(\kappa_{xx} - \kappa_{xx}^{ph})$ (Fig. 3B). The observed thermal-Hall angle is orders of magnitude smaller than the prediction (26), which calls for further studies. It is also an open question why continuous excitations are observed in the present compound whereas they are absent in κ -(BEDT-TTF)₂Cu₂(CN)₃ (18). Such excitations may be strongly suppressed by inhomogeneity.

References and Notes

- P. A. Lee, *Science* **321**, 1306 (2008).
- P. W. Anderson, *Science* **235**, 1196 (1987).
- Y. Shimizu, K. Miyagawa, K. Kanoda, M. Maesato, G. Saito, *Phys. Rev. Lett.* **91**, 107001 (2003).
- T. Itou, A. Oyamada, S. Maegawa, M. Tamura, R. Kato, *J. Phys. Conf. Ser.* **145**, 012039 (2009).
- T. Itou, A. Oyamada, S. Maegawa, M. Tamura, R. Kato, *Phys. Rev. B* **77**, 104413 (2008).
- D. A. Huse, V. Elser V, *Phys. Rev. Lett.* **60**, 2531 (1988).
- O. I. Motrunich, *Phys. Rev. B* **72**, 045105 (2005).
- S.-S. Lee, P. A. Lee, *Phys. Rev. Lett.* **95**, 036403 (2005).
- X.-G. Wen, *Phys. Rev. B* **65**, 165113 (2002).
- D. N. Sheng, O. I. Motrunich, M. P. A. Fisher, *Phys. Rev. B* **79**, 205112 (2009).
- G. Misguich, C. Lhuillier, B. Bernu, C. Waldtmann, *Phys. Rev. B* **60**, 1064 (1999).
- Y. Qi, C. Xu, S. Sachdev, *Phys. Rev. Lett.* **102**, 176401 (2009).
- V. Kalmeyer V, R. B. Laughlin, *Phys. Rev. Lett.* **59**, 2095 (1987).
- H. Morita, S. Watanabe, M. Imada, *J. Phys. Soc. Jpn.* **71**, 2109 (2002).
- T. Yoshioka, A. Koga, N. Kawakami, *Phys. Rev. Lett.* **103**, 036401 (2009).
- S. Yunoki, S. Sorella, *Phys. Rev. B* **74**, 014408 (2006).
- Y. Hayashi, M. Ogata, *J. Phys. Soc. Jpn.* **76**, 053705 (2007).
- M. Yamashita *et al.*, *Nat. Phys.* **5**, 44 (2009).
- L. D. Fadddeev, L. A. Takhtajan, *Phys. Lett.* **85A**, 375 (1981).
- F. D. M. Haldane, *Phys. Lett.* **93A**, 464 (1983).
- S. Yamashita *et al.*, *Nat. Phys.* **4**, 459 (2008).
- Y. Shimizu, K. Miyagawa, K. Kanoda, M. Maesato, G. Saito, *Phys. Rev. B* **73**, 140407 (2006).
- R. S. Manna, M. de Souza, A. Brühl, J. A. Schlueter, M. Lang, *Phys. Rev. Lett.* **104**, 016403 (2010).
- M. Tamura, R. Kato, *Polyhedron* **24**, 2817 (2005).
- A. V. Sologubenko, K. Giannò, H. R. Ott, A. Vietkine, A. Revcolevschi, *Phys. Rev. B* **64**, 054412 (2001).
- Materials and methods are available as supporting material on Science Online.
- R. Moessner, S. L. Sondhi, *Phys. Rev. Lett.* **86**, 1881 (2001).
- Ch. Waldtmann *et al.*, *Eur. Phys. J. B* **2**, 501 (1998).
- T. Grover, N. Trivedi, T. Senthil, P. A. Lee, "Weak Mott insulators on the triangular lattice: Possibility of a gapless nematic quantum spin liquid;" <http://arxiv.org/abs/0907.1710> (2009).
- H. Katsura, N. Nagaosa, P. A. Lee, *Phys. Rev. Lett.* **104**, 066403 (2010).
- We thank M. P. A. Fisher, S. Fujimoto, T. Itou, H. Katsura, N. Kawakami, H. Kawamura, P. A. Lee, G. Misguich, Y. Nakazawa, and A. Tanaka for valuable discussion. This work was supported by Grants-in-Aid for Scientific Research (KAKENHI) from the Japan Society for the Promotion of Science and the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), a Grant-in-Aid for the Global Centers of Excellence Program "The Next Generation of Physics, Spun from Universality and Emergence," and a Grant-in-Aid for Scientific Research on Innovative Areas "Heavy Electrons" (No. 20102006) from MEXT, Japan.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5983/1246/DC1
Materials and Methods
Figs. S1 and S2
Table S1
References

10 February 2010; accepted 29 April 2010
10.1126/science.1188200

Collective Lamb Shift in Single-Photon Superradiance

Ralf Röhlsberger,^{1,*} Kai Schlage,¹ Balaram Sahoo,¹ Sebastien Couet,² Rudolf Ruffer³

Superradiance, the cooperative spontaneous emission of photons from an ensemble of identical atoms, provides valuable insights into the many-body physics of photons and atoms. We show that an ensemble of resonant atoms embedded in the center of a planar cavity can be collectively excited by synchrotron radiation into a purely superradiant state. The collective coupling of the atoms via the radiation field leads to a substantial radiative shift of the transition energy, the collective Lamb shift. We simultaneously measured the temporal evolution of the superradiant decay and the collective Lamb shift of resonant ⁵⁷Fe nuclei excited with 14.4-kilo-electron volt synchrotron radiation. Our experimental technique provides a simple method for spectroscopic analysis of the superradiant emission.

The development of quantum electrodynamics is very closely related to the discovery and explanation of the Lamb shift of atomic energy levels (1, 2). The Lamb shift is a small

energy shift of bound atomic states, mainly resulting from the emission and reabsorption of virtual photons within the same atom. An additional contribution emerges if many identical

two-level atoms are interacting collectively with a resonant radiation field. In this case, a virtual photon that is emitted by one atom may be reabsorbed by another atom within the ensemble. The resulting collective Lamb shift scales with the optical density of the atoms and sensitively depends on their spatial arrangement (3–9). At high atomic densities, however, atom-atom interactions mask the collective Lamb shift, making it almost impossible to observe. Since the early theoretical studies (3), only one measurement of a collective line shift has been reported for a multiphoton excitation scheme in a noble gas (10, 11). Experimental assessment of the collective Lamb

¹Deutsches Elektronen Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany. ²Instituut voor Kern- en Stralingsfysica and INPAC, Katholieke Universiteit Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium. ³European Synchrotron Radiation Facility, B.P. 220, 38043 Grenoble Cedex, France.

*To whom correspondence should be addressed. E-mail: ralf.roehlsberger@desy.de

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
<http://www.sciencemag.org/cgi/content/full/328/5983/1248>

Supporting Online Material can be found at:
<http://www.sciencemag.org/cgi/content/full/science.1187770/DC1>

This article **cites 26 articles**, 1 of which can be accessed for free:
<http://www.sciencemag.org/cgi/content/full/328/5983/1248#otherarticles>

This article has been **cited by 1 articles** hosted by HighWire Press; see:
<http://www.sciencemag.org/cgi/content/full/328/5983/1248#otherarticles>

This article appears in the following **subject collections**:
Physics
<http://www.sciencemag.org/cgi/collection/physics>

shift for single-photon excitation, particularly in solid-state samples, has remained elusive. Here, we measured the collective Lamb shift for an ensemble of ^{57}Fe Mössbauer nuclei (transition energy $E_0 = \hbar\omega_0 = 14.4$ keV, level width $\Gamma_0 = 4.7$ neV, natural lifetime $\tau_0 = \hbar/\Gamma_0 = 141$ ns, where ω_0 is the frequency and $\hbar$ is the Planck constant divided by 2π), embedded in a planar cavity that was resonantly excited by synchrotron radiation x-rays.

The collective Lamb shift is a cooperative optical effect that is intimately connected with the phenomenon of superradiance—the cooperative spontaneous emission of radiation from an ensemble of identical two-level atoms, introduced by Dicke in 1954 (12) and observed experimentally after short-pulse lasers became available (13, 14). Superradiance was originally conceived by Dicke for sample dimensions R much smaller than the wavelength $\lambda = 2\pi c/\omega_0$ of the resonant transition (where c is the speed of light). In this limit, an ensemble of N atoms, excited by a single photon, decays collectively N times as quickly as a single isolated atom. It is, however, the opposite limit $R \gg \lambda$ that is relevant for most experiments in the optical or x-ray regimes. For an extended sample, the state vector of the atomic ensemble is given by

$$|\Psi_+\rangle_{k_0} = \frac{1}{\sqrt{N}} \sum_j \exp(i\mathbf{k}_0 \cdot \mathbf{r}_j) |b_1 b_2 \dots a_j \dots b_N\rangle \quad (1)$$

where $|b_1 b_2 \dots a_j \dots b_N\rangle$ is a Fock state in which the j th atom at position $\mathbf{r}_j$ is in the excited state a and all other atoms are in the ground state b . Because the positions $\mathbf{r}_j$ translate to different times of excitation t_j for pulsed directional excitation with wave vector $\mathbf{k}_0$, $t_j = \mathbf{k}_0 \cdot \mathbf{r}_j / \omega_0$, this state has been called the “timed” Dicke state (15), closely resembling the “nuclear exciton” state for ensembles of Mössbauer nuclei collectively excited by synchrotron radiation (16). The spatial phase factors of the timed Dicke state lead to directional emission along $\mathbf{k}_0$ (15, 16), with the temporal evolution of its probability amplitude given by $\beta_+(t) = \beta_0 \exp[-(1/2)(\Gamma_0 + \Gamma_N + iL_N)t/\hbar]$, where Γ_0 is the single-atom decay width, Γ_N is the N -atom decay width, and L_N is the collective Lamb shift. This shows that the emission from extended samples can exhibit Dicke-like superradiant dynamics as in the limit $R \ll \lambda$ (5, 6). In this case, the collective Lamb shift is not masked by atom-atom interactions as in the small-sample limit, thus becoming experimentally accessible.

Note that the preparation of the timed Dicke state requires the excitation probability to be uniform over the sample (17). In conventional forward scattering, however, the timed Dicke state can never be created exactly because the radiation from atoms at the entrance front of the sample interferes destructively with radiation from atoms deeper in the sample (16). The result is a nonuniformly excited ensemble of atoms. With increasing time after excitation, a substantial frac-

tion of nonsuperradiant, slowly decaying states is populated (16, 18). The interference of these states results in temporal oscillations that are often referred to as dynamical beats or propagation quantum beats (19). To avoid the population of nonsuperradiant states, it is necessary for the sample to be optically thin upon absorption and optically thick upon emission in order to exhibit strong superradiant enhancement (20).

We prepare the timed Dicke state of Eq. 1 by embedding an ensemble of ^{57}Fe atoms in a planar low- Q cavity ($Q < 100$) and resonantly exciting them with synchrotron radiation pulses coupled evanescently into the first-order mode (Fig. 1). The cavity geometry exhibits a number of important features that facilitate the observation of the collective Lamb shift. First, the ensemble appears to be optically thin upon absorption, with the excitation probability for all atoms within the ensemble being practically the same. Second, the partial waves emitted from the ensemble are funneled into a single mode of the cavity, yielding a total amplitude corresponding to the emission from an optically thick sample. Therefore, the optical thicknesses of the sample for absorption and emission are effectively decoupled. Finally, the cavity enhances the photonic density of states at the position of the atoms, which strongly amplifies the radiative level shift, as has been shown for single atoms in cavities (21).

In a typical cavity used here (Fig. 1), a material of low electron density (C, 32 nm thick) is sandwiched between two layers of high electron

density (Pt). The top Pt layer is thin enough (2.2 nm) that x-rays impinging under grazing angles can evanescently couple into the cavity. Guided modes are excited at angles of incidence where deep minima in the x-ray reflectivity appear (Fig. 1B). An ensemble of ^{57}Fe nuclei is embedded as ultrathin layer within the guiding layer of the cavity. The spectral response of this system (i.e., its reflectivity R in the vicinity of the nuclear resonance) is calculated by introducing the nuclear forward scattering amplitude f_N of the resonant layer material. If we assume that the nuclei within the ensemble exhibit a single Lorentzian resonance line, then f_N is given by

$$f_N(x) = \frac{2\pi c_N \rho_N}{k_0 k_{0z}} \left[\frac{1}{x + i} \right] \quad (2)$$

where $k_{0z} \approx k_0 \varphi$ is the component of the wave vector $\mathbf{k}_0$ normal to the surface, ρ_N is the number density of resonant nuclei, c_N combines nuclear parameters (22), and $x = \Delta E/(\Gamma_0/2)$ describes the detuning $\Delta E = E - E_0$ of the photon energy from the exact resonance energy E_0 . We perform a perturbation expansion of the nuclear resonant reflectivity R of the cavity in powers of f_N . A diagrammatical representation of the perturbation series of R (Fig. 1A) shows that the amplitudes A_i of the partial waves emitted from the nuclear ensemble at the “vertices” (black dots) correspond to the various orders of the perturbation expansion. The scattered amplitude in the n th outgoing wave is related to the $(n-1)$ th amplitude via $A_n =$

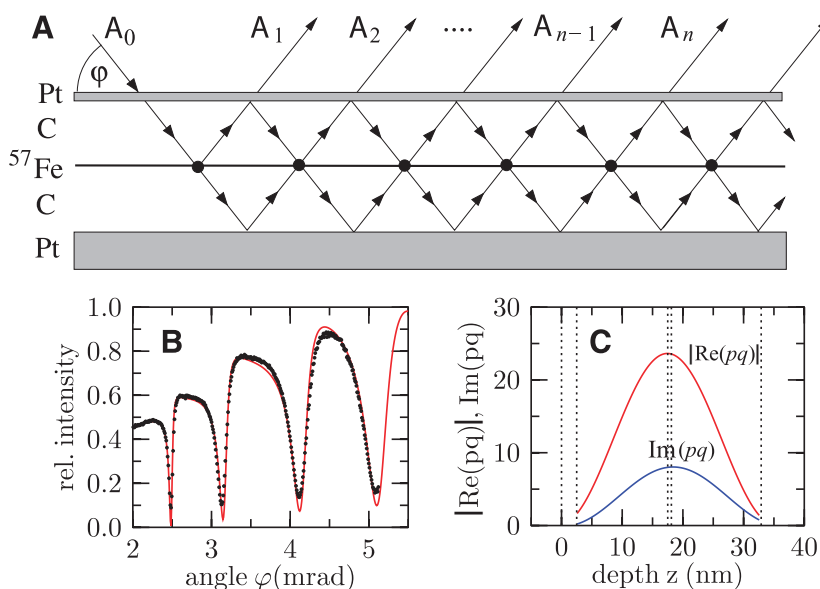


Fig. 1. (A) Structure of the planar cavity and scattering geometry used for calculation of the collective Lamb shift for the resonant ^{57}Fe nuclei embedded in the planar cavity. (B) Measured (nonresonant) x-ray reflectivity of one of the samples (sample 1) used in the experiment, consisting of (2.2 nm Pt)/(16 nm C)/(0.6 nm ^{57}Fe)/(16 nm C)/(13 nm Pt) deposited on a superpolished Si substrate with a root-mean-square roughness below 0.3 nm. The solid line is a fit to the data, from which the exact values of the layer thicknesses were determined. Guided modes are excited at the angular positions of the minima. (C) The quantities $\text{Im}(pq)$ and $|\text{Re}(pq)|$ as function of depth in the first-order mode (23). Dashed lines mark the interfaces between layers.

$(idf_N)pqA_{n-1}$, where d is the thickness of the ^{57}Fe layer and p and q are the amplitudes of the wave fields (at the position of the resonant nuclei) propagating in the directions of the incident and the reflected beams, respectively. p and q can be cal-

culated via a propagation matrix formalism (23). The depth dependence of the relevant product pq for the first-order mode of the cavity used here is shown in Fig. 1C. For the first vertex, we have $A_1 = (idf_N)p^2A_0$, which also includes the coupling of

the radiation into the cavity. Thus, the sum over all orders results in

$$R = idf_N p^2 \sum_{k=0}^{\infty} (idpqf_N)^k = \frac{idp^2 f_N}{1 - idpqf_N} \quad (3)$$

Inserting $f_N(x) = C/(x + i)$, where $C = 2\pi c_N \rho_N d / (k_0 k_{0z})$ and $x = \Delta E / (\Gamma_0/2)$, we obtain again a Lorentzian resonance line $R(\Delta E) = Cp^2(\Gamma_0/2)[\Delta E + L_N + i(\Gamma_0 + \Gamma_N)/2]$ that exhibits a collective decay width of $\Gamma_N = C\text{Re}(pq)\Gamma_0 = \chi\Gamma_0$ (where χ is the collective enhancement factor) and an energy shift of $L_N = -C[\text{Im}(pq)\Gamma_0/2]$. Because the factor C scales with the density ρ_N of the resonant nuclei, the energy shift is essentially the collective Lamb shift of the ensemble of nuclei in the cavity. Explicitly, the collective Lamb shift is given by

$$L_N = -\frac{1}{2} \left[\frac{2\pi c_N \rho_N d}{k_0^2} \frac{d}{\varphi} \Gamma_0 \right] \text{Im}(pq) \quad (4)$$

The expression in square brackets is the collective decay width for the free-standing layer of resonant nuclei illuminated under an angle φ . The factor $\text{Im}(pq)$ describes the enhancement due to the cavity. In the first-order mode, $\text{Im}(pq)$ peaks in the center of the guiding layer (Fig. 1C). For a nuclear ensemble placed at that depth, one can expect a radiative level shift up to $-13\Gamma_0$, depending on the layer materials the cavity consists of. This value is eventually limited by the electronic photoabsorption of the Fe atoms, which disturbs the wave field in the cavity.

To observe these effects, we prepared cavity samples containing ultrathin ^{57}Fe layers with thicknesses of 0.6 nm (sample 1) and 1.2 nm (sample 2) by physical vapor deposition techniques (23). After pulsed resonant excitation at $t = 0$ with synchrotron radiation (23), we observed for both samples an exponential decay $I(t) = I_0 \exp[-(\Gamma_0 + \Gamma_N)t/\hbar] = I_0 \exp[-(1 + \chi)t/\tau_0]$ over two orders of magnitude from $t = 6$ ns until $t = 20$ ns (Fig. 2). χ describes the speedup of the decay with values of $\chi = 42$ and $\chi = 61$ for samples 1 and 2, respectively.

After 20 ns, the temporal response levels off into a much slower decay, caused by inhomogeneous broadening and a residual quadrupole hyperfine interaction of the nuclei in the sample. The inset of Fig. 2 shows a conversion electron Mössbauer (CEMS) spectrum of ^{57}Fe (0.6 nm) embedded in a carbon matrix from which a quadrupole splitting (QS) of 33.8 neV, an isomer shift of 4.7 neV, and a QS distribution of 37.6 neV (full width at half maximum) was derived. Taking into account the measured hyperfine interaction parameters, we used the dynamical theory of nuclear resonant scattering (16, 24–26) to calculate the temporal evolution of the nuclear decay and obtained very good agreement with the measured data. As the width of the inhomogeneously broadened resonance line is much smaller than the collective decay width Γ_N , the inhomogeneous dephasing has negligible effects on the superradiant decay (23, 27, 28). Thus, the observation of the initial,

Fig. 2. Nuclear resonant time response of the two samples upon excitation of the first-order guided mode in sample 1 (0.6 nm ^{57}Fe , upper curve) and sample 2 (1.2 nm ^{57}Fe , lower curve). Both curves are displaced along the ordinate for better visualization. The solid lines are theoretical calculations taking into account the hyperfine interaction of the ^{57}Fe nuclei as derived from the CEMS spectrum shown in the inset. From the initial slope of both curves, speedup values of $\chi = 42$ and $\chi = 61$, respectively, are determined. The dotted line at the top of the figure displays the natural decay of the ^{57}Fe .

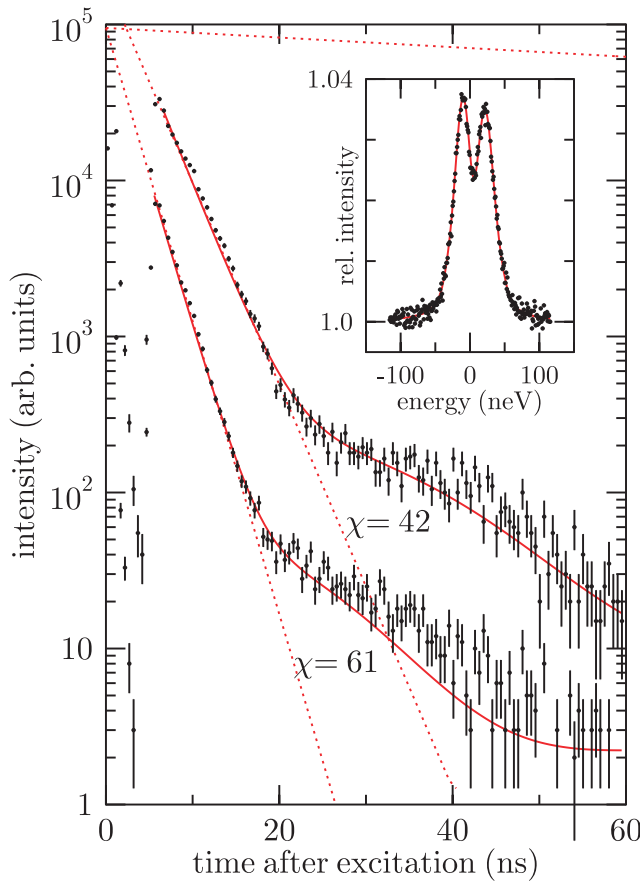
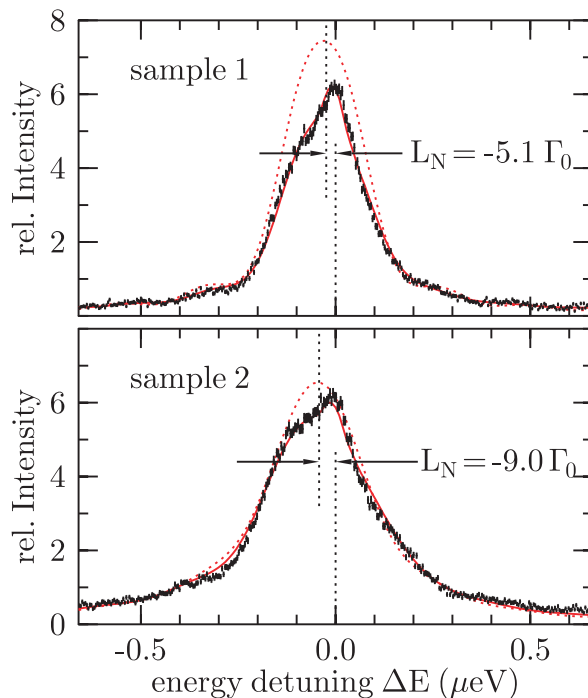


Fig. 3. Energy response of the two samples, as recorded using a stainless steel foil (thickness 5.6 μm) as an analyzer. Delayed quanta were collected in a time window between 22 ns and 160 ns. The shift of the center of mass of these curves relative to the origin corresponds to the collective Lamb shift. The collective decay widths are $\Gamma_N = 45\Gamma_0$ and $\Gamma_N = 65\Gamma_0$, respectively. Solid red curves are theoretical calculations according to (23). For comparison, the dashed red lines are calculations assuming vanishing hyperfine interaction.



strictly exponential decay in Fig. 2 proves the formation of the timed Dicke state in Eq. 1 and its superradiant decay.

According to Eq. 4, we expect a collective Lamb shift of several natural linewidths Γ_0 for both samples. To measure this shift, we analyzed the energy spectra $|R(E)|^2$ of the radiation reflected from the samples. For that purpose, we used a Doppler drive (as used in conventional Mössbauer spectroscopy) on which a thin foil (thickness 5.6 μm) consisting of stainless steel with the composition $^{57}\text{Fe}_{55}\text{Cr}_{25}\text{Ni}_{20}$ was mounted in transmission geometry. In this alloy the ^{57}Fe nuclei exhibit a single-line resonance. While the foil was moved back and forth in a constant-acceleration mode [i.e., with a linear variation of the Doppler shift $E_D = (v_D/c)E_0$], delayed resonant quanta within a time interval (t_1 , t_2) were counted as function of E_D (23). Figure 3 shows the measured energy spectra of both samples being aligned to their first-order guided modes. In both cases, the center of mass of the measured spectra was shifted toward negative energies with values of $-5.1\Gamma_0$ for sample 1 and $-9.0\Gamma_0$ for sample 2. These values are reproduced by Eq. 4 within a margin of about $\pm 5\%$, which confirms that the collective Lamb shift in this approximation is proportional to the extension (thickness) of the nuclear ensemble. The solid lines in Fig. 3 are theoretical simulations (23), indicating a good agreement with the measured spectra. The asymmetric shape of the energy spectra is a consequence of the residual quadrupole hyperfine interaction to which the ^{57}Fe nuclei are subjected (29). For comparison, the dashed lines in Fig. 3 are calculations with vanishing quadrupole interaction, indicating that the measurement of the radiative level shift is not affected by the hyperfine interaction (except for an isomer shift) of the nuclei in the sample.

Under grazing angles, the relevant component of the wave vector is the one normal to the surface, $k_{0z} \approx \varphi k_0$, and therefore $k_{0z}R \ll 1$. Thus, all nuclei within the thin layer experience the same amplitude and phase in the standing wave pattern of the cavity mode while being confined to a dimension that is small relative to the field wavelength $\lambda_z = 2\pi/k_{0z}$. As a result, the relative phases in Eq. 1 are equal modulo 2π , so that effectively the small-sample limit $k_0R \ll 1$ of Dicke superradiance is realized. In this limit the sample effectively appears to be two-dimensional, where the collective Lamb shift scales with the areal density ρ_A of the nuclei as $L_N \sim \rho_A \lambda^2$. This is in contrast to three-dimensional samples such as a spherical cloud, for which the collective Lamb shift scales as $L_N \sim \rho \lambda^3$ (3, 5). Because ρ_A can be easily adjusted via the layer thickness, the planar cavity appears to be an attractive system for studying cooperative optical phenomena.

References and Notes

- W. E. Lamb Jr., R. C. Retherford, *Phys. Rev.* **72**, 241 (1947).
- H. A. Bethe, *Phys. Rev.* **72**, 339 (1947).
- R. Friedberg, S. R. Hartmann, J. T. Manassah, *Phys. Rev. C* **7**, 101 (1973).
- A. Svidzinsky, J. T. Chang, *Phys. Rev. A* **77**, 043833 (2008).
- M. O. Scully, *Phys. Rev. Lett.* **102**, 143601 (2009).
- M. O. Scully, A. A. Svidzinsky, *Science* **325**, 1510 (2009).
- R. Friedberg, J. T. Manassah, *Phys. Lett. A* **373**, 3423 (2009).
- J. T. Manassah, *Laser Phys.* **20**, 259 (2010).
- R. Friedberg, J. T. Manassah, *Phys. Lett. A* **374**, 1648 (2010).
- W. R. Garrett, R. C. Hart, J. E. Wray, I. Datskou, M. G. Payne, *Phys. Rev. Lett.* **64**, 1717 (1990).
- R. Friedberg, S. R. Hartmann, J. T. Manassah, *Phys. Rev. A* **39**, 93 (1989).
- R. H. Dicke, *Phys. Rev.* **93**, 99 (1954).
- N. Skribanowitz, I. P. Herman, J. C. Mac Gillivray, M. S. Feld, *Phys. Rev. Lett.* **30**, 309 (1973).
- M. Gross, S. Haroche, *Phys. Rep.* **93**, 301 (1982).
- M. O. Scully, E. S. Fry, C. H. R. Ooi, K. Wódkiewicz, *Phys. Rev. Lett.* **96**, 010501 (2006).
- J. P. Hannon, G. T. Trammell, *Hyperfine Interact.* **123–124**, 127 (1999).
- M. O. Scully, *Laser Phys.* **17**, 635 (2007).
- I. E. Mazets, G. Kurizki, *J. Phys. B* **40**, F105 (2007).
- U. van Bürck, *Hyperfine Interact.* **123–124**, 483 (1999).
- B. W. Adams, *J. Mod. Opt.* **56**, 1974 (2009).
- D. J. Heinzen, M. S. Feld, *Phys. Rev. Lett.* **59**, 2623 (1987).
- $c_N = \{f_{LM}/[2(1 + \alpha)]\}[(2I_g + 1)/(2I_e + 1)]$, where f_{LM} is the Lamb-Mössbauer factor, α is the coefficient of internal conversion, and I_g and I_e are the nuclear spins of the ground and excited state, respectively. For metallic ^{57}Fe at room temperature, $f_{LM} \approx 0.8$, so that $c_N \approx 0.09$.
- See supporting material on Science Online.
- W. Sturhahn, E. Gerdau, *Phys. Rev. B* **49**, 9285 (1994).
- R. Röhlberger, *Hyperfine Interact.* **123–124**, 301 (1999).
- W. Sturhahn, *Hyperfine Interact.* **125**, 149 (2000).
- C. Leonardi, A. Vaglica, *Nuovo Cim. B* **67**, 256 (1982).
- V. V. Temnov, U. Woggon, *Phys. Rev. Lett.* **95**, 243602 (2005).
- Because the energetic width of the cavity resonance (~ 100 eV) is much larger than the width of the nuclear response (~ 300 neV), any impact of the cavity resonance on the nuclear spectral shape can be excluded (weak-coupling limit of cavity QED).
- We thank A. N. Poddubny for valuable discussions and for helpful comments on the manuscript; H. Lipkin, J. T. Manassah, B. W. Adams, and K. Temst for fruitful discussions on this and related topics; and W. Pfützner for assistance during sample preparation. Supported by the Geconcerteerde Onderzoeksacties and Fonds Wetenschappelijk Onderzoek Flanders programs (S.C.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1187770/DC1
Materials and Methods
Figs. S1 to S3
References

1 February 2010; accepted 13 April 2010
Published online 13 May 2010;
10.1126/science.1187770
Include this information when citing this paper.

Dynamic Kinetic Resolution of Biaryl Atropisomers via Peptide-Catalyzed Asymmetric Bromination

Jeffrey L. Gustafson, Daniel Lim, Scott J. Miller*

Despite the widespread use of axially chiral, or atropisomeric, biaryl ligands in modern synthesis and the occurrence of numerous natural products exhibiting axial chirality, few catalytic methods have emerged for the direct asymmetric preparation of this compound class. Here, we present a tripeptide-derived small-molecule catalyst for the dynamic kinetic resolution of racemic biaryl substrates. The reaction proceeds via an atropisomer-selective electrophilic aromatic substitution reaction using simple bromination reagents. The result is an enantioselective synthesis that delivers chiral nonracemic biaryl compounds with excellent optical purity and good isolated chemical yields (in most cases a $>95:5$ enantiomer ratio and isolated yields of 65 to 87%). A mechanistic model is advanced that accounts for the basis of selectivity observed.

The stereochemical implications of hindered rotation in nonplanar molecules, termed atropisomerism, have intrigued chemists for at least 89 years (1, 2). Atropisomeric

compounds exhibit an axis of chirality (Fig. 1A), rather than a stereogenic atom, such as an sp^3 -hybridized carbon with four distinct substituents (Fig. 1B). The capacity of the single bond be-

tween two aromatic rings to freely rotate is the basis of racemization for many atropisomeric compounds (3). Yet in naturally occurring compounds, atropisomeric molecules are often found in single isomeric form because of substituents on aryl rings that raise the barriers to racemization. Also, the localization of aryl rings within multicyclic ring systems can constrain single bond rotations, preventing isomerization and the observation of mixtures (4, 5). These properties no doubt contribute to the remarkable structures and functions of numerous biologically active compounds that contain single atropisomers as part of their structure (6). Perhaps the glycopeptide antibiotic vancomycin is the signature bioactive natural product of this type (Fig. 1C). The chiral ligand BINAP [2,2'-bis(diphenylphosphino)-1,1'-binaphthyl] (Fig. 1D), a venerable ligand for

Department of Chemistry, Yale University, 225 Prospect Street, Post Office Box 208107, New Haven, CT 06520-8107, USA.

*To whom correspondence should be addressed. E-mail: scott.miller@yale.edu

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1251>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1251/DC1>

This article **cites 33 articles**, 1 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/328/5983/1251#otherarticles>

This article appears in the following **subject collections**:

Chemistry

<http://www.sciencemag.org/cgi/collection/chemistry>

enantioselective catalysis, may be the best-known designed example (7).

Despite the prevalence and importance of atropisomerism in organic structures, the field of asymmetric catalysis has not yet recorded extensive success in the development of catalysts that control this stereochemical feature. In several excellent cases, asymmetric metal-catalyzed formation of the biaryl bond has resulted in high enantioselectivity for certain substrates (8–12). Pioneering work by Bringmann (13) and Clayden (14) focused instead on stoichiometrically inducing the selective reaction of a single atropisomer in a dynamic mixture of freely rotating, rapidly racemizing biaryls. Catalytic reactions of this nature are presently rare, and only modest atropisomer selectivity has been observed (15). Here, we report a simple chiral catalyst that mediates highly enantioselective electrophilic aromatic substitution reactions, thereby promoting atropisomer-selective functionalization of rapidly racemizing biaryl compounds.

We began our project by identifying a substrate class that would exhibit axial chirality with atropisomers that might rapidly interconvert. Compound **1** fulfills this criterion, with a barrier to atropisomer interconversion that may be estimated to be ~7 kcal/mol (16). Upon reaction with *N*-bromosuccinimide (NBS) in the presence of catalytic *N*-methylpyrrolidine [10 mole % (mol %) **2**], compound **1** was triply brominated to yield **3** (Scheme 1), which exhibits much more restricted rotation about the bond connecting the aromatic rings. In fact, the barrier to racemization for compound **3** may be estimated to be ~30 kcal/mol (17). This barrier is sufficiently high so that in principle, if an enantioselective catalyst could be found for the conversion of **1** to **3**, isolation of optically enriched, nonracemizing products could be obtained at room temperature. Notably, singly or doubly brominated intermediates were not observed under the reaction conditions that we used.

We then turned our attention to the issue of enantioselectivity. Our choice of catalyst for these studies followed an empirical path, given the paucity of precedents for this type of asymmetric catalytic reaction. Nonetheless, we were guided by the principle that Lewis base catalysis of electrophilic bromination reactions is possible (18–23). Furthermore, we were driven by the recognition that simple peptide-based catalysts have substantial capacity to mediate a wide variety of mechanistically diverse enantioselective transformations (24, 25). Moreover, we had recently shown that peptide-based catalysts exhibit remarkable enantioselectivity in the derivatization of unusual aromatic compounds (26).

We chose peptide **4** as a starting place for catalyst screening (Fig. 2), with the rationale that the chiral environment of the peptide β turn (27) could lead to atropisomer selection during the electrophilic aromatic substitution reaction. β -*N,N*-Dimethylamino alanine (Dmaa) was introduced as the *N*-terminal residue with hope that it might

function as an additional basic site, favoring catalyst-substrate contacts. This first-generation catalyst provided encouraging results, with methyl ester (**1a**) giving the product in nonracemic form (entry 1). Swapping **1a** for benzyl amide (**1b**) improved enantioselectivity to nearly 2:1 (entry 2). Nitro group substitution led to lower selectivity (entry 3). However, carboxylic acid (**1d**) proved a very promising substrate for the reaction, affording **3d** with a 75:25 enantiomeric ratio (e.r.) (entry 4). We were particularly encouraged by the mathematical implications of this result for dynamic kinetic resolution (28). A 75:25 ratio of enantiomers at high levels of conversion requires that substrate racemization occur (29), at least to some extent, during the course of the overall reaction. Moreover, we observed no erosion of e.r.s upon extended storage or heating (100°C, 15 hours), consistent with the expected high barriers to biaryl bond rotation/

racemization required for our fundamental premise to operate.

We then turned our attention to the optimization of the catalyst structure. Tripeptide catalysts were initially examined with the Dmaa residue at the *N* terminus and chiral α -methylbenzylamides (α Mba) at the *C* terminus (Fig. 3). Simple replacement of L-Pro in **4** for L-pipecolinic acid (Pip) led to a notable improvement in selectivity (catalyst **5**, entry 2). Substitution of a range of amino acids in the *i* + 2 position reduced enantioselectivity slightly (entries 3 to 5), whereas altering the stereogenic sense at this position (entry 6) or the adjacent *C*-terminal position (entry 7) led to an even greater loss in selectivity. Even so, catalyst **11a**, with a simplified *N,N*-dimethyl amide at the *C* terminus, delivered **3d** in the highest observed enantioselectivity under the initial reaction conditions (entry 8). Ester substitution in place of the *N,N*-dimethyl amide lowered selec-

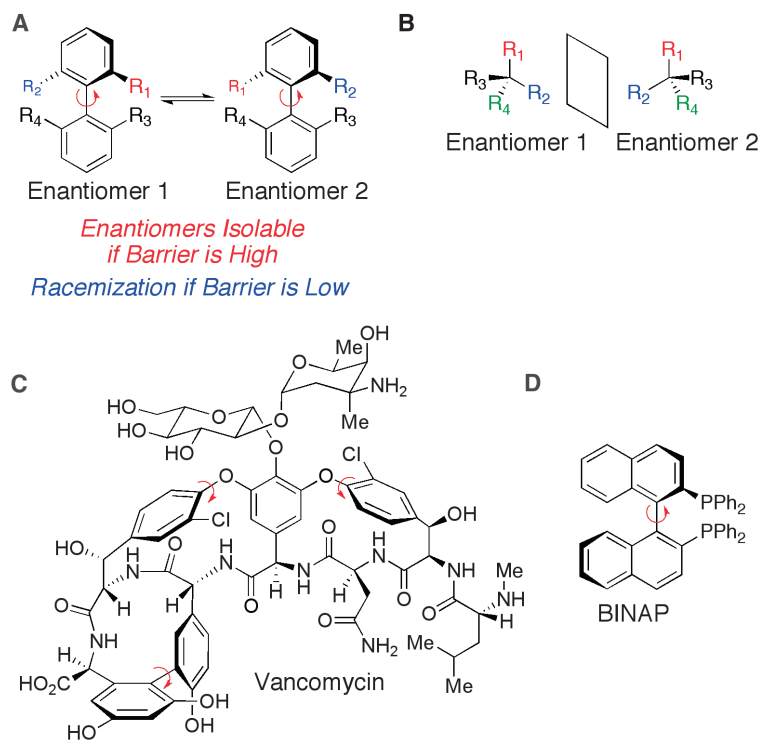
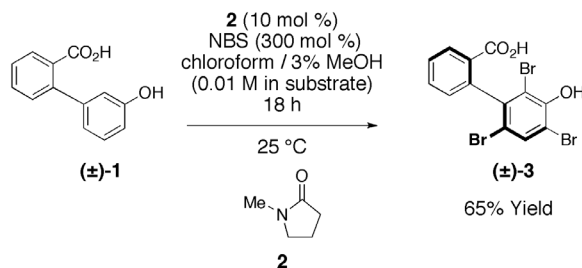
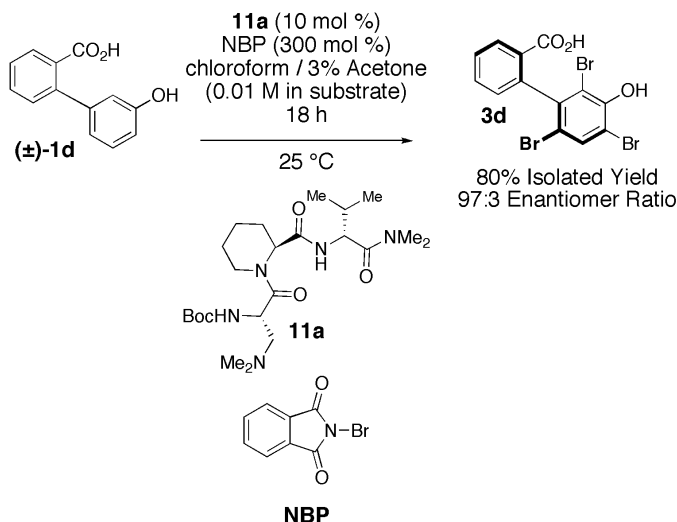


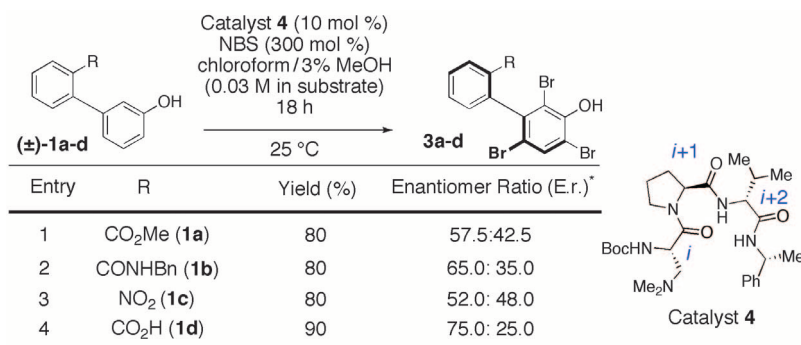
Fig. 1. (A) Biaryl atropisomers are isolable if the barrier to rotation about the single bond linking the rings is high. The enantiomers interconvert via racemization if the barrier is low. R groups are listed in Fig. 2. (B) *sp*³-hybridized carbon atoms with four different substituents form generally stable enantiomers. (C) Vancomycin is a natural product that exists as a single atropisomer. Me, methyl group. (D) BINAP [2,2'-bis(diphenylphosphino)-1,1'-binaphthyl] is a widely used chiral ligand. Ph, phenyl group.



Scheme 1.



Scheme 2.



^{*}The major atropisomer of **3d** was assigned to the *R*-configuration by X-ray analysis.

Fig. 2. Initial screen of catalysis for asymmetric bromination of **1**. Isolated yields correspond to the anisole methyl ester after treatment with 4 equivalents of trimethylsilyl (TMS)–diazomethane (2 M in Et₂O) for 15 min in 0.2 M toluene:MeOH (3:1). This work-up assists in purification and e.r. determination by chiral high performance liquid chromatography (HPLC). See (30) for details. Enantiometer ratios were determined by chiral HPLC.

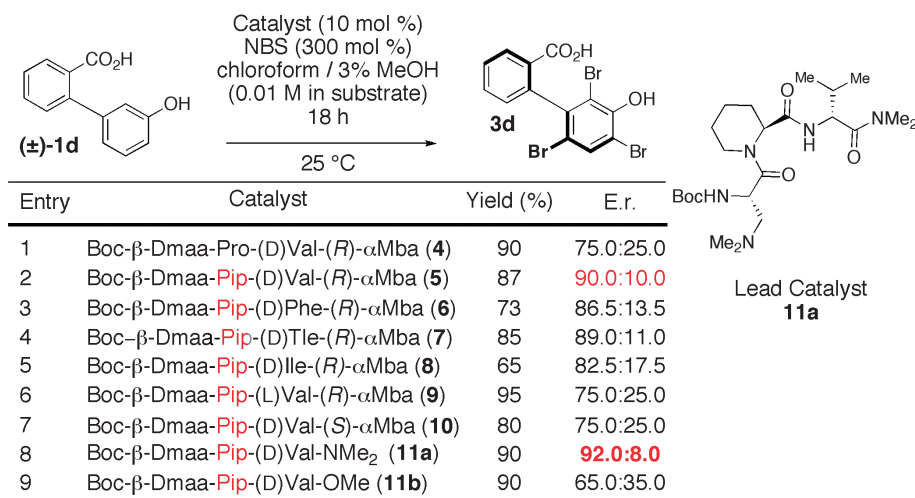


Fig. 3. Optimization of the catalyst structure. Data determined as in Fig. 2. Boc, *tert*-butoxycarbonyl.

tivity substantially (entry 9). We therefore declared peptide **11a** our lead catalyst for further study.

We then evaluated additional reaction parameters, including solvents, concentration, temperature, and brominating agent. Among these factors, the bromine source proved particularly influential. *N*-Bromophthalimide (NBP) consistently gave the best results. In combination with optimized solvent and concentration conditions (30), catalyst **11a** (10 mol %) was found to promote conversion of racemic **(±)-1d** to an 80% yield of **3d**, with an e.r. of 97:3 (Scheme 2). These results were obtained at room temperature on a 0.5 mmol scale, and we observed no variation over a range of reaction scales up to 9 mmol (~2 g).

With an effective catalyst and mild reaction conditions in hand, we then examined substrate scope. The substrates shown in Fig. 4 were readily prepared by using established aryl-aryl bond-forming cross-coupling reactions (31). A range of substituents on the aryl rings of the substrate proved compatible with the excellent results obtained for substrate **1d**. For example, electron-withdrawing NO₂-group substitution para to the phenolic arene, as in compound **12**, led to the isolation of **13** with a 97:3 e.r. (entry 2). Compound **14**, with *meta*-NO₂ substitution, gave a similar result (entry 3), as did electron-rich anisoles **16** (entry 4) and **18** (entry 5). Given the importance of fluorinated aromatic compounds, we assessed *meta*-substituted fluorobenzene (**20**) and *para*-substituted fluorinated substrate **22** (32) and found that they also were excellent substrates for the reaction (entries 6 and 7). Alkyl substitution ortho to the carboxylic acid leads to somewhat lower enantioselectivity (entry 8).

It appears that this catalytic methodology could be quite useful for asymmetric synthesis of a range of heteroarene compounds containing functionalities of relevance to bioactive natural product substructures (33). For example, pyrrole analog **26** was converted to **27** with a moderate but encouraging 85:15 e.r. (entry 9). Catechol **28**, a potential substructure in the stegane natural products (34), was converted to **29** with an e.r. of 95:5 (entry 10), and can be enriched to exhibit e.r.s of >98:2 by a single recrystallization. The results shown in Fig. 4, taken together, reveal that catalyst **11a** and, perhaps more importantly, its analogs could be quite useful conceptually for synthesis of a broad range of optically enriched biaryl-type compounds.

We have also gained some insight into the mechanism of these intriguing reactions. Several experiments with stripped down, potentially catalytic moieties confirm the possibility of amide catalysis (20), perhaps via a type of [O-Br]-cationic species. For example, when the reaction is conducted in the absence of a catalyst, bromination is sluggish, and **1** is converted to **3** in only 15% yield after 18 hours (Fig. 5). Tertiary amines such as *N,N*-diisopropylethylamine also provide only slight rate acceleration, and the yield of **3** is 30% under analogous conditions. However, the use of

Entry	Racemic Starting Material	Product	Yield (%)	E.r.	Entry	Racemic Starting Material	Product	Yield (%)	E.r.
1			80	97.0:3.0	6			70	97.0:3.0
2			85	97.0:3.0	7			65	96.5:3.5
3			75	96.5:3.5	8			85	87.0:13.0
4			70	96.0:4.0	9			77*	85.0:15.0
5			80	94.0:6.0	10			70	95.0:5.0

* 400 mol % of NBP.

Fig. 4. Substrate scope for enantioselective bromination. Data determined as in Fig. 2. Results represent the average of two to three runs per substrate.

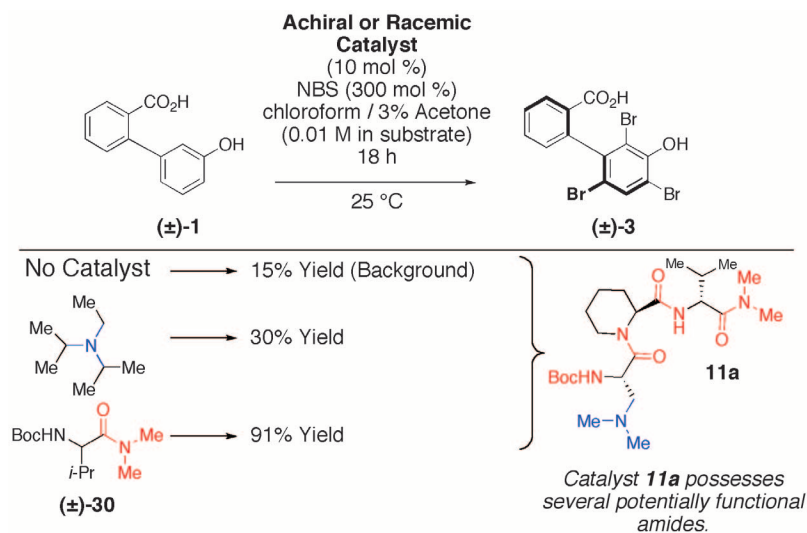
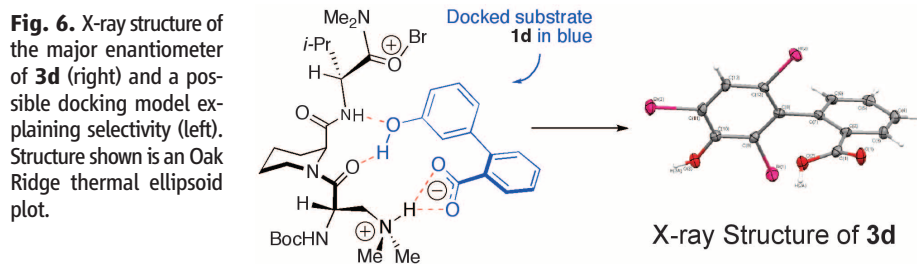


Fig. 5. Assessment of the catalytic efficiency of simple functional groups. *i*-Pr, *iso*-propyl group.



the *N,N*-dimethylamide **30** leads to 91% isolated yield of **3**. These results point to a functional role for one of the several amides resident in catalyst **11a**, including the terminal *N,N*-dimethylamide. Neither mono- nor dibrominated products were observed in substantial quantities under these conditions.

These observations, in combination with our knowledge of stereochemical aspects of the reaction, have allowed us to posit a possible explanation for the stereochemical outcome. Single-crystal, heavy-atom x-ray analysis allowed assignment of the absolute configuration of the major product of enantioenriched **3d** as the (*R*)-atropisomer (Fig. 6). An accessible conformation of catalyst **11a** is likely as shown in Fig. 6, with axial disposition of the pipecolinic acid substituent, because of the well-known preference of *N*-acyl piperidines to adopt conformations with axial 2-substituents to avoid allylic strain (35). Docking of the substrate **1d** through salt bridge formation between the Dmaa tertiary amine and the substrate carboxylic acid disposes a putative *O*-bromonium ion toward formation of the observed stereoisomer of **3d**. Notably, free rotation of partially brominated (e.g., mono- and dibrominated) species may be possible until the ortho-ortho' substituents are each installed, leading to a barrier to rotation high enough to preclude product racemization. On the other hand, hydrogen bonds between the phenolic proton and the catalyst amide groups may prevent such bond rotation at intermediate stages of the reaction. Although other mechanisms may also be operative, models

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1255>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1255/DC1>

This article **cites 25 articles**, 9 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/328/5983/1255#otherarticles>

This article appears in the following **subject collections**:

Chemistry

<http://www.sciencemag.org/cgi/collection/chemistry>

such as that offered in Fig. 6 provide one starting place for evaluation.

Synthesis of optically enriched biaryl compounds using enantioselective catalysts and dynamic kinetic resolution should enable improved access to stereodefined atropisomeric materials. More broadly, the approach described herein may also stimulate related research involving selective reactions of other interconverting, axially chiral compounds, promoted by simple peptide-based catalysts.

References and Notes

- J. Kenner, W. V. A. Stubbings, *J. Chem. Soc.* **119**, 593 (1921).
- R. Adams, H. C. Yuan, *Chem. Rev.* **12**, 261 (1933).
- R. W. Stoughton, R. Adams, *J. Am. Chem. Soc.* **54**, 4426 (1933).
- E. L. Eiel, S. H. Wilen, Eds., *Stereochemistry of Organic Compounds* (Wiley, New York, 1994), pp. 1142–1148.
- D. Seebach *et al.*, *Tetrahedron* **53**, 7539 (1997).
- J. Clayden, W. J. Moran, P. J. Edwards, S. R. LaPlante, *Angew. Chem. Int. Ed.* **48**, 6398 (2009).
- R. Noyori, H. Takaya, *Acc. Chem. Res.* **23**, 345 (1990).
- G. Bringmann *et al.*, *Angew. Chem. Int. Ed.* **44**, 5384 (2005).
- M. Shindo, K. Koga, K. Tomioka, *J. Am. Chem. Soc.* **114**, 8732 (1992).
- X. Li, J. B. Hewgley, C. A. Mulrooney, J. Yang, M. C. Kozlowski, *J. Org. Chem.* **68**, 5500 (2003).
- J. Yin, S. L. Buchwald, *J. Am. Chem. Soc.* **122**, 12051 (2000).
- T. Hayashi, K. Hayashizaki, T. Kiyoi, Y. Ito, *J. Am. Chem. Soc.* **110**, 8153 (1988).
- G. Bringmann, D. Menche, *Acc. Chem. Res.* **34**, 615 (2001).
- J. Clayden, S. P. Fletcher, J. J. W. McDouall, S. J. M. Rowbottom, *J. Am. Chem. Soc.* **131**, 5331 (2009).
- F. Kakiuchi, P. Le Gendre, A. Yamada, H. Ohtaki, S. Murai, *Tetrahedron Asymmetry* **11**, 2647 (2000).
- G. Bott, L. D. Field, S. Sternhell, *J. Am. Chem. Soc.* **102**, 5618 (1980).
- A. I. Meyers, R. J. Himmelsbach, *J. Am. Chem. Soc.* **107**, 682 (1985).
- S. E. Denmark, G. L. Beutner, *Angew. Chem. Int. Ed.* **47**, 1560 (2008).
- W. Zhang, H. Xy, H. Xu, W. J. Tang, *J. Am. Chem. Soc.* **131**, 3832 (2009).
- R. H. Mitchell, Y. H. Hai, R. V. Williams, *J. Org. Chem.* **44**, 4733 (1979).
- S. France, A. Weatherwax, T. Lectka, *Eur. J. Org. Chem.* **2005**, 475 (2005).
- A. Sakakura, A. Ukai, K. Ishihara, *Nature* **445**, 900 (2007).
- S. Ahmad, D. C. Braddock, G. Hermitage, *Tetrahedron Lett.* **48**, 915 (2007).
- S. J. Miller, *Acc. Chem. Res.* **37**, 601 (2004).
- E. Colby Davie, S. M. Mennen, Y. Xu, S. J. Miller, *Chem. Rev.* **107**, 5759 (2007).
- C. A. Lewis *et al.*, *J. Am. Chem. Soc.* **130**, 16358 (2008).
- T. S. Haque, J. C. Little, S. H. Gellman, *J. Am. Chem. Soc.* **118**, 6975 (1996).
- R. Noyori, M. Tokunaga, M. Kitamura, *Bull. Chem. Soc. Jpn.* **68**, 36 (1995).
- J. Steinreiber, K. Faber, H. Griengl, *Chem. Eur. J.* **14**, 8060 (2008).
- Experimental procedures and compound characterization are available as supporting material on Science Online.
- A. F. Litke, C. Dai, G. C. Fu, *J. Am. Chem. Soc.* **122**, 4020 (2000).
- D. A. Watson *et al.*, *Science* **325**, 1661 (2009).
- R. Martin *et al.*, *Angew. Chem. Int. Ed.* **48**, 8042 (2009).
- A. S. Kende, L. S. Liebeskind, *J. Am. Chem. Soc.* **98**, 267 (1976).
- F. Johnson, *Chem. Rev.* **68**, 375 (1968).
- We are grateful to the National Institute of General Medical Sciences of the NIH for support (GM068649). Metrical parameters for the solid state structure of compound **3d** are available free of charge from the Cambridge Crystallographic Data Centre under CCDC-773872.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5983/1251/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S10
References and Notes

16 February 2010; accepted 21 April 2010
10.1126/science.1188403

Operation Mechanism of a Molecular Machine Revealed Using Time-Resolved Vibrational Spectroscopy

Matthijs R. Panman,¹ Pavol Bodis,¹ Daniel J. Shaw,¹ Bert H. Bakker,¹ Arthur C. Newton,¹ Euan R. Kay,² Albert M. Brouwer,^{1*} Wybren Jan Buma,^{1*} David A. Leigh,^{2*} Sander Woutersen^{1*}

Rotaxanes comprise macrocycles that can shuttle between docking stations along an axle. We explored the nanosecond shuttling mechanism by reversing the relative binding affinities of two stations through ultraviolet-induced transient reduction. We monitored the ensuing changes in the CO-stretching bands of the two stations and the shuttling macrocycle by means of an infrared probing pulse. Because hydrogen-bond scission and formation at the initial and final stations led to well-resolved changes in the respective CO-stretch frequencies, the departure and arrival of the macrocycle could be observed separately. We found that the shuttling involves two steps: thermally driven escape from the initial station, followed by rapid motion along the track ending either at the initial or final station. By varying the track's length, we found that the rapid motion approximates a biased one-dimensional random walk. However, surprisingly, the direction of the overall motion is opposite that of the bias.

Recent developments in the generation and control of motion on a molecular scale have enabled the construction of prototypical molecular machines (1–14). The mechanics of such molecular machines are fundamentally different from those of macroscopic machines. As a machine's size and speed approach those of the solvent molecules surrounding it, macroscopic

concepts such as viscous friction lose meaning, and thermal fluctuations lead to intrinsic randomness in the machine's behavior (15, 16). A detailed understanding of the motion of molecular machines requires experiments with both structural sensitivity at the atomic level and sufficient time resolution. Time-resolved vibrational spectroscopy meets both of these requirements (17, 18). Probing the stretch modes of specific, localized chemical bonds affords a direct view of the conformational changes underlying the operation of individual components of the molecular machine. The time resolution of the probing is determined by the pulse duration, which can be less than 100 fs. Here, we used time-resolved

vibrational pump-probe experiments on a light-triggered rotaxane shuttle to unravel its operating mechanism.

Rotaxanes consist of mechanically interlocked wheel- and axle-like components. Figure 1 shows the chemical structure of this type of molecular machine, together with its operation cycle, which we characterized previously with spectroelectrochemistry and transient ultraviolet-visible (UV-Vis) spectroscopy (3, 19). In the neutral molecule, the macrocycle is hydrogen-bonded predominantly (>99%) to the succinamide station (**succ**) (3, 19). After excitation of the naphthalimide station (**ni**) with a 355-nm light pulse, the rotaxane undergoes rapid (~1.6 ns) intersystem crossing to the triplet state (20). In this state, the **ni** station is reduced by an external electron donor to form a radical anion (**ni^{•-}**). The **ni^{•-}** station has a much greater affinity for the macrocycle than does the **succ** station [equilibrium constant >1000 (3)]. Consequently, the macrocycle shuttles over the thread and forms hydrogen bonds to the **ni^{•-}** station. After slow (~100 μs) charge recombination between **ni^{•-}** and the radical cation of the electron donor, the macrocycle travels back over the thread and binds to the **succ** station, and the system is ready to shuttle again (3). In the following, rotaxanes with track length *n* are referred to as C_{*n*}.

The vibrational absorption spectrum of the C₁₂ rotaxane in its initial state is shown in Fig. 2A. All peaks in the rotaxane spectrum can be assigned by comparison with spectra of the constituent components (19). The symmetric and antisymmetric CO-stretch modes of the **ni** station are observed as peaks 1 and 7, respectively (the latter also contains a contribution from the CO-stretch vibration

¹Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, Netherlands. ²School of Chemistry, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JJ, UK.

*To whom correspondence should be addressed. E-mail: a.m.brouwer@uva.nl (A.M.B.); w.j.buma@uva.nl (W.J.B.); david.leigh@ed.ac.uk (D.A.L.); s.woutersen@uva.nl (S.W.)

of the macrocycle), and the CO-stretch mode of the **succ** station is observed as peak 5.

In the time-resolved experiments, we trigger the rotaxane with a 4-ns, 355-nm pulse and observe the subsequent vibrational absorption change by means of a 100-fs mid-infrared (IR) probe pulse (21, 22). In Fig. 2, B and C, we show normalized (19, 21) UV-IR transient spectra of the C₁₂ rotaxane at different delay times. The spectra at early delays (<120 ns) show how the triplet state evolves to the radical anion state (**ni**^{•−}). Subsequently, the spectrum evolves further as the macrocycle shuttles from the **succ** to the **ni**^{•−} station (Fig. 2C). For the assignment of the peaks in Fig. 2C we use steady-state spectra of the initial, charged, and final states (19). Peaks 2 and 3 (labeled ARRIVAL) are the CO-stretch modes of the **ni**^{•−} station when it is free and when it is hydrogen-bonded to the macrocycle, respectively. The temporal evolution of these peaks therefore directly mirrors the arrival of the macrocycle at the **ni**^{•−} station. Similarly, peaks 4 and 5 (labeled DEPARTURE) are the CO-stretch modes of the **succ** station when free and when hydrogen-bonded to the macrocycle. The behavior of these two peaks therefore mirrors the departure of the macrocycle from the **succ** station. Peak 6 is due to the CO-stretch mode of the macrocycle when its NH groups are hydrogen-bonded to the **ni**^{•−} station (19). The corresponding absorption decrease of the CO-stretch mode of **succ**-bound macrocycles as they leave the **succ** station is superimposed on (temporally static) peak 7.

The time evolution of the intensity of peaks 2, 3, 4, 5, and 6 is shown in Fig. 3. We find that the departure and arrival of the macrocycle can be well described with the same time constant. We performed a global least-squares fit to the data (21), in which the charging and shuttling are each described by a single rate constant. From the fit (see curves in Fig. 3) we obtain $k_{\text{charge}} = 2.8 (\pm 0.4) \times 10^7 \text{ s}^{-1}$, which corresponds to the charging lifetime of $35 \pm 5 \text{ ns}$. We find that for the C₁₂ rotaxane, $k_s^{\text{C}_{12}} = 1.30 (\pm 0.03) \times 10^6 \text{ s}^{-1}$; this shuttling rate constant corresponds to a shuttling time of $0.77 \pm 0.02 \mu\text{s}$, in agreement with previous UV-Vis experiments (3).

The fact that the departure and arrival exhibit the same dynamics implies that during the shuttling process, an individual macrocycle spends only a short time on the thread relative to the average shuttling time of the ensemble. The observed average shuttling time is therefore not determined by the time the macrocycle spends on the CH₂ chain, but rather by the time it takes for the macrocycle to escape from the initial station. The sequence of events in the shuttling of this rotaxane is thus as follows: occasional escape from the **succ** station, followed by a rapid motion over the thread, ending at either the **ni**^{•−} (where it can no longer escape because of the stronger hydrogen bonding) or the initial **succ** station (where it will wait until the next escape). Below, we investigate the two steps of this mechanism in more detail.

To quantify the energy barrier that the macrocycle must overcome to escape from the **succ** station, we measured the shuttling rate of a C₉ rotaxane (23) over a temperature range of 45 K (21). The temperature increase due to the partial conversion of the pump energy into heat is negligible (21). The temperature dependence of $k_s^{\text{C}_9}$ displays Arrhenius behavior, indicating that the escape from the **succ** station is effectively a single-barrier event and that the energy required by the macrocycle to cross it is provided by thermal fluctuations. We find that the data can be well described with the Eyring equation,

$$k_s(T) = \frac{k_B T}{h} \exp \left[- \left(\frac{\Delta G^\ddagger}{RT} \right) \right] \quad (1)$$

where k_B is the Boltzmann factor, R is the gas constant, h is the Planck constant, $\Delta G^\ddagger$ is the Gibbs free energy of activation, and T is the absolute temperature of the system. From a least-squares fit (fig. S5), we find that the enthalpy of activation $\Delta H^\ddagger$ is $26 (\pm 1) \text{ kJ mol}^{-1}$ and the entropy of activation $\Delta S^\ddagger$ is $-32 (\pm 3) \text{ J mol}^{-1} \text{ K}^{-1}$. The energy barrier $\Delta H^\ddagger$ is consistent with the breaking of approximately four hydrogen bonds, assuming a binding energy of $-7.5 (\pm 0.8) \text{ kJ mol}^{-1}$ per hydrogen bond (24). The negative activation entropy suggests that

some ordering of the system is required for escaping from the **succ** station. Folding of the thread onto itself can be excluded, as this should be evident in the **ni** and/or **ni**^{•−} CO-stretch response in the transient spectra (19). The ordering might involve a relative positioning of macrocycle and thread that is favorable for breaking the macrocycle-**succ** hydrogen bonds.

The next step in the shuttling process is the fast motion of the macrocycle over the thread, ending at either the **succ** station or the **ni**^{•−} station. To investigate the nature of this fast motion, we explored how changing the carbon-chain length affects the probability of the macrocycle ending at the **ni**^{•−} station. This probability is proportional to the observed shuttling rate, as the rate at which escape from the **succ** station occurs is independent of the carbon-chain length. The shuttling rates of rotaxanes with thread lengths $n = 5, 9, 12$, and 16 are shown in Fig. 4. The rate of shuttling was observed to decrease markedly with increasing thread length. We analyzed the data by modeling the motion of the macrocycle over the thread as a one-dimensional random walk. In the simplest version of this model, the C_{*n*} chain is modeled as a track of n local free-energy minima, between which the macrocycle makes random jumps, to end up at either the **ni**^{•−} or **succ** station. If we assume equal probabilities for the macrocycle to make one step toward

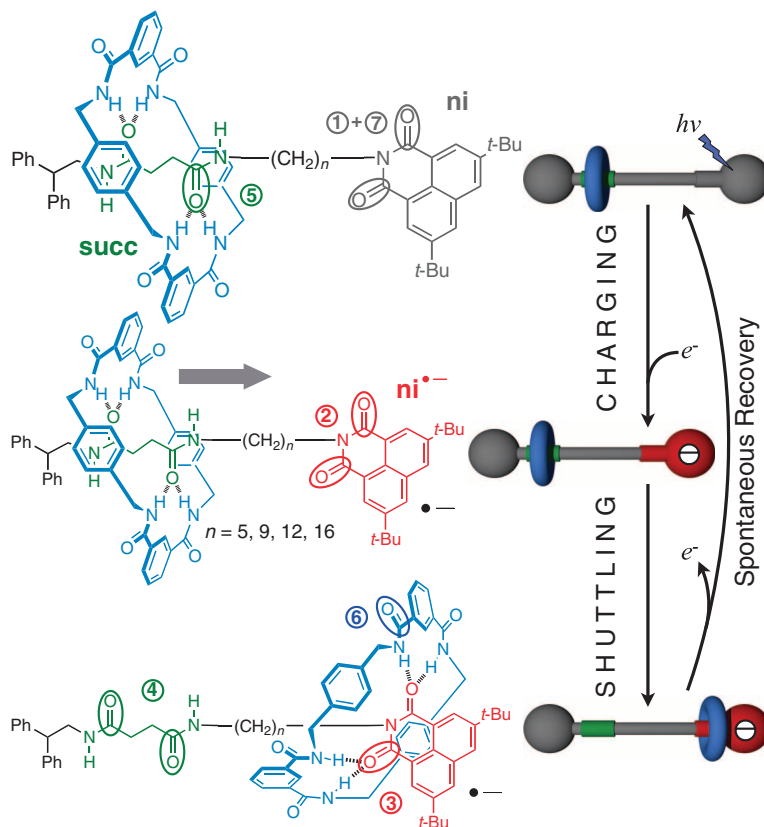


Fig. 1. Chemical structures of the [2]rotaxane shuttle in the neutral, radical anion, and shuttled radical anion states. 1,4-Diazabicyclo[2.2.2]octane (DABCO) was used as the external electron donor. We studied derivatives with different track lengths n . The labeled CO groups correspond to the labeled peaks in Fig. 2. Ph, phenyl; Bu, butyl.

either station, the probability of the macrocycle arriving at the ni^+ station is given by

$$P_s(n) = \frac{1}{n+1} \quad (2)$$

Fig. 2. (A) Fourier-transform IR spectrum of 10^{-4} M C_{12} rotaxane and 10^{-2} M DABCO in CD_3CN ; path length 11 mm (solvent subtracted). (B) Transient UV-IR spectra of the C_{12} rotaxane at delays ranging from 10 to 120 ns after UV excitation. (C) Transient UV-IR spectra of the C_{12} rotaxane at delays ranging from 120 to 2000 ns after UV excitation. The labeled vibrations in the transient spectrum correspond to the labeled CO groups of Fig. 1.

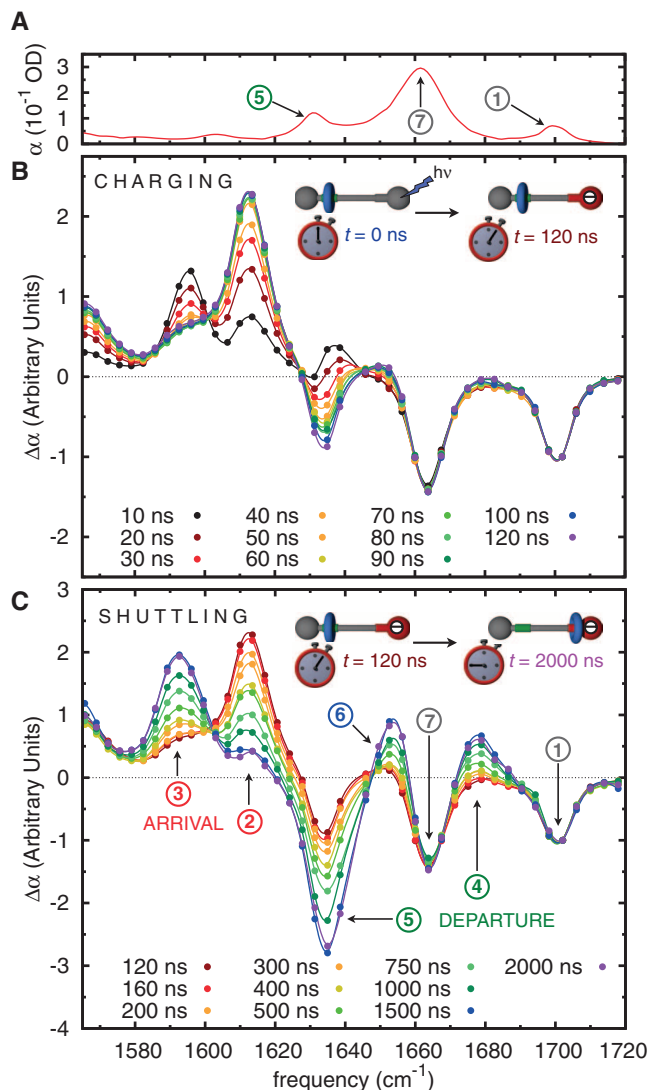
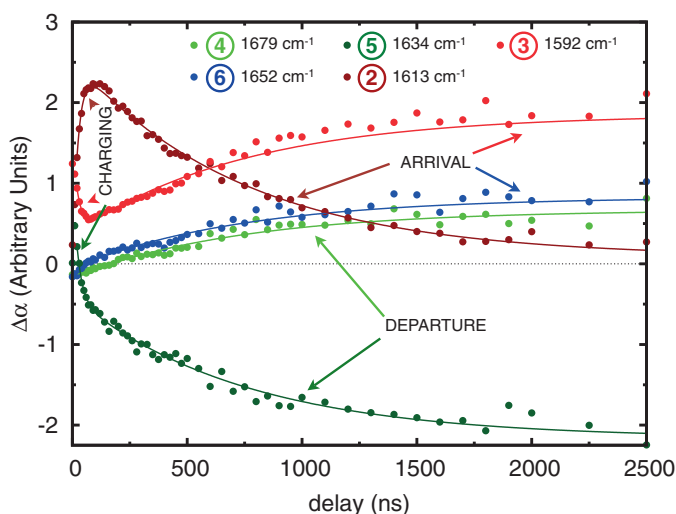


Fig. 3. Observed time dependence of the absorption peaks 2, 3, 4, 5, and 6 normalized to the maximum intensity of peak 1, each of which corresponds to a specific CO bond in a specific state of the operation cycle; see Fig. 1. The curves represent least-squares fits of bi-exponential decays.



(25). A least-squares fit of the rate data to this equation, with the proportionality factor between the observed rate constant and $P_s(n)$ as the only fit parameter, is shown as the gray dashed curve in Fig. 4. Although the model reproduces the ob-

served trend, the quantitative agreement is poor. Allowing for a certain number of CH_2 units to be occupied by the macrocycle (25) still results in a poor description of the data (21).

We find that very good agreement is obtained by introducing a small bias in the probabilities of making a step in the forward or backward direction. Assuming a probability p of hopping forward (in the direction of the ni^+ station), and $1 - p$ of hopping backward, the probability of arriving at the ni^+ station becomes

$$P_s(n) = \frac{1 - \left(\frac{1-p}{p}\right)}{1 - \left[\left(\frac{1-p}{p}\right)^{n+1}\right]} \quad (3)$$

(26). From a least-squares fit (with p and an overall scaling factor as the only fit parameters), we find that the best description of the data is obtained for $p = 0.442 \pm 0.003$ (red curve in Fig. 4). This value of p implies that the event of the macrocycle moving one step toward the ni^+ station is slightly less probable than the event of it moving toward the succ station ($1 - p = 0.56$); that is, the random translational motion of the macrocycle along the track has a small bias toward the succ station. At present, we can only speculate as to the driving force behind this bias. An enthalpic driving force seems improbable in view of the short-range nature of the hydrogen-bond interactions that give rise to the overall ΔH of the shuttling. It is more likely that the bias is mainly of entropic origin. This could be the case if the thread conformation is on average slightly more favorable to macrocycle translation at the succ side than at the ni^+ side. Interestingly, the overall motion of the macrocycle from succ to ni^+ occurs against the bias (which is toward the succ station). This is possible only because there is a global free-energy minimum at the ni^+ station.

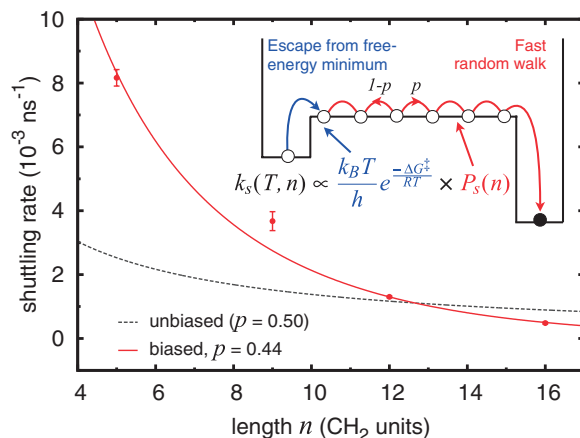
Our findings can be summarized by combining Eqs. 1 and 3 to obtain an expression for the shuttling rate as a function of temperature and track length:

$$k_s(T, n) \propto \frac{k_B T}{h} \exp\left[-\left(\frac{\Delta G^\ddagger}{RT}\right)\right] P_s(n) \quad (4)$$

Each term describes one stage of the shuttling mechanism (Fig. 4, inset). The probability $P_s(n)$ for the random walk to end at the final station is $\ll 1$ [for the shortest shuttle, $P_s(5) = 0.09$; for the longest, $P_s(16) = 0.005$], so that many escapes (tens to hundreds) of the macrocycle from the initial station are required before a successful shuttling event occurs. This leads to a certain amount of randomness in the timing of the arrival of the shuttle.

A more detailed understanding of the bias against shuttling requires an exploration of the rotaxane's free-energy landscape using molecular dynamics simulations. Because shuttling is such a rare event, this will require state-of-the-art simulation methods, and we hope that our results will stimulate work in this direction. On a more fundamental level, it is noteworthy that the

Fig. 4. Thread length (n) dependence on the shuttling rate constant k_s fitted with an unbiased random walk model and a biased random walk model (Eq. 3). The error bars represent $\pm 1\sigma$ (21). The inset shows a schematic representation of the shuttling mechanism.



unpredictability of the arrival time of the shuttle does not imply unpredictability of the arrival itself. A consideration of the probability for a macrocycle to make the transition from one station to another is essentially an attempt to describe the shuttling motion—about which our experiments provide ensemble thermodynamic and kinetic information—at the single-molecule level. Fluctuation theorems (27) provide a connection between thermodynamic parameters and single-molecule measurements of molecular mechanical events (28). In particular, theoretical work on time-symmetry breaking in such measurements has shown that the free-energy change per operation cycle should be about 4 to 8 $k_B T$ for an individual molecular machine to rectify thermal fluctuations to such an extent as to advance mostly forward in time (i.e., toward the conformation corresponding to the free-energy minimum) (29). The free-energy difference that drives the rotaxane's shuttling motion (the difference between the **succ**- and **ni**⁺-bound macrocycle) is -18 kJ/mol (3), which corresponds to 7.3 $k_B T$. This is just enough to ensure that the

shuttling occurs in the forward time direction (29). The rotaxane-based shuttles are thus reliable in the sense that the macrocycle will eventually arrive at the **ni**⁺ station. On the other hand, our results show that with increasing length of the shuttle's track, the time at which this arrival occurs becomes more and more unpredictable.

References and Notes

1. T. R. Kelly, H. De Silva, R. A. Silva, *Nature* **401**, 150 (1999).
2. N. Koumura, R. W. J. Zijlstra, R. A. van Delden, N. Harada, B. L. Feringa, *Nature* **401**, 152 (1999).
3. A. M. Brouwer *et al.*, *Science* **291**, 2124 (2001).
4. D. A. Leigh, J. K. Y. Wong, F. Dehez, F. Zerbetto, *Nature* **424**, 174 (2003).
5. P. Thordarson, E. J. A. Bijsterveld, A. E. Rowan, R. J. M. Nolte, *Nature* **424**, 915 (2003).
6. J. D. Badjic, V. Balzani, A. Credi, S. Silvi, J. F. Stoddart, *Science* **303**, 1845 (2004).
7. J. V. Hernández, E. R. Kay, D. A. Leigh, *Science* **306**, 1532 (2004).
8. S. P. Fletcher, F. Dumur, M. M. Pollard, B. L. Feringa, *Science* **310**, 80 (2005).
9. V. Balzani *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1178 (2006).
10. T. Muraoka, K. Kinbara, T. Aida, *Nature* **440**, 512 (2006).
11. J. E. Green *et al.*, *Nature* **445**, 414 (2007).

12. V. Serreli, C. F. Lee, E. R. Kay, D. A. Leigh, *Nature* **445**, 523 (2007).
13. S. Hiraoka, E. Okuno, T. Tanaka, M. Shiro, M. Shionoya, *J. Am. Chem. Soc.* **130**, 9089 (2008).
14. M. von Delius, E. M. Geertsema, D. A. Leigh, *Nat. Chem.* **2**, 96 (2010).
15. R. D. Astumian, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1843 (2005).
16. R. D. Astumian, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19715 (2007).
17. A. Stolow, D. M. Jonas, *Science* **305**, 1575 (2004).
18. C. Kolano, J. Helbing, M. Kozinski, W. Sander, P. Hamm, *Nature* **444**, 469 (2006).
19. D. C. Jagesar, F. Hartl, W. J. Buma, A. M. Brouwer, *Chem. Eur. J.* **14**, 1935 (2008).
20. D. C. Jagesar *et al.*, *Adv. Funct. Mater.* **19**, 3440 (2009).
21. See supporting material on Science Online.
22. The time resolution of the experiment is determined by the pump pulse, which is 4 ns. The 100-fs probe pulse provides us with the spectral bandwidth necessary to probe the entire rotaxane absorption spectrum in one laser pulse.
23. We chose the C₉ rotaxane for the temperature studies because of its superior solubility in CD₃CN and therefore its better signal-to-noise ratio relative to the C₁₂ rotaxane.
24. M. Akiyama, T. Ohtani, *Spectrochim. Acta A* **50**, 317 (1994).
25. A. B. C. Deutman *et al.*, *Science* **322**, 1668 (2008).
26. G. Grimmett, D. Welsh, *Probability: An Introduction* (Oxford Univ. Press, Oxford, 1986).
27. C. Jarzynski, *Phys. Rev. E* **73**, 046105 (2006).
28. D. Collin *et al.*, *Nature* **437**, 231 (2005).
29. E. H. Feng, G. E. Crooks, *Phys. Rev. Lett.* **101**, 090602 (2008).
30. We thank P. Reinders, S. Kettelarij, D. Bebelaar, M. Groeneveld, H. Luiten, H. Beukers, and C. van den Biggelaar for technical support, and G. Rothenberg for valuable discussions. Supported by the European Community (EMMA IHP Research Training Network, contract HPRN-CT- 2002-00168, and Hy3M STREP, contract NMP4-CT-2004-013525) and by the Stichting voor Fundamenteel Onderzoek der Materie (FOM), which is financially supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5983/1255/DC1
Materials and Methods
Figs. S1 to S7
References

4 February 2010; accepted 19 April 2010
10.1126/science.1187967

The Thermodynamics of the Elusive HO₃ Radical

Sébastien D. Le Picard,^{1*} Meryem Tizniti,¹ André Canosa,¹ Ian R. Sims,^{1*} Ian W. M. Smith^{2*}

The role of HO₃ as a temporary reservoir of atmospheric OH radicals remains an open question largely because of the considerable uncertainty in the value of the dissociation energy of the HO–O₂ bond (D_0) or, equivalently, the standard enthalpy of formation of HO₃ ($\Delta_f H^\circ$). Using a supersonic flow apparatus, we have observed by means of laser-induced fluorescence the decay of OH radicals in the presence of O₂ at temperatures between 55.7 and 110.8 kelvin (K). Between 87.4 and 99.8 K, the OH concentration approached a nonzero value at long times, allowing equilibrium constants for the reaction with O₂ to be calculated. Using expressions for the equilibrium constant from classical and statistical thermodynamics, and values of partition functions and standard entropies calculated from spectroscopic data, we derived values of $D_0 = (12.3 \pm 0.3)$ kilojoules per mole and $\Delta_f H^\circ (298 \text{ K}) = (19.3 \pm 0.5)$ kilojoules per mole. The atmospheric implications of HO₃ formation are therefore very slight.

The weakly bound HO₃ radical has been postulated as a transient intermediate in three processes of importance in the Earth's

atmosphere (Scheme 1, reactions R1 to R3). Sridharan *et al.* (1) showed that when ¹⁸O was used as the atomic reactant in the rapid reaction

R1 with HO₂, the hydroxyl product was exclusively ¹⁶OH, suggesting that the reaction occurred not through H-atom abstraction but rather via transient formation of H¹⁶O¹⁶O¹⁸. Vibrationally excited OH formed in reaction R2 is the source of the Meinel bands emitted from the upper atmosphere (2). The reaction leads to high yields of the OH product in the highest energetically accessible vibrational levels $v = 7$ to 9 (3, 4), indicating energy release on a highly attractive potential energy surface (5), which would be the form expected if HO₃ is weakly bound relative to OH + O₂ (4). A third process—R3, the relaxation

¹Institut de Physique de Rennes, Equipe Astrochimie Expérimentale, UMR 6251 du CNRS, Bâtiment 11c, Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France.
²The University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

*To whom correspondence should be addressed. E-mail: sebastien.le-picard@univ-rennes1.fr (S.D.L.); ian.sims@univ-rennes1.fr (I.R.S.); iwms2@cam.ac.uk (I.W.M.S.)

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1258>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1258/DC1>

This article **cites 19 articles**, 2 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/328/5983/1258#otherarticles>

This article appears in the following **subject collections**:

Chemistry

<http://www.sciencemag.org/cgi/collection/chemistry>

of vibrationally excited $\text{OH}(\nu = 1)$ by O_2 —is rapid relative to the rates found for similar, nonradical collision partners, and this observation led McCabe *et al.* (6) to postulate that relaxation is facilitated by moderately strong attractive forces between OH and O_2 .

However, the most important effect of HO_3 in the Earth's atmosphere would arise if it could serve as a temporary reservoir for OH radicals after their association with O_2 . Validating this possibility requires the accurate measurement or calculation of equilibrium constants for reaction R4 at the temperatures prevailing in the atmosphere.

In 2005, Suma *et al.* (7) recorded the rotational spectrum of HO_3 created in a supersonic jet. Analyzing this spectrum, they obtained the rotational constants for the trans-conformer of HO_3 , from which the rotational partition function and the rotational contribution to the standard

entropy for HO_3 at different temperatures can be estimated. However, obtaining a value for D_0 (the dissociation energy of the $\text{HO}-\text{O}_2$ bond) or $\Delta_f H^\circ$ (the standard enthalpy change for reaction R4) has proved difficult, and the value of this quantity is crucial in determining whether HO_3 has an appreciable atmospheric role.

HO_3 was unequivocally detected in experiments based on neutralization-reionization and neutralization-reionization/collisionally activated mass spectrometry, by using HO_3^+ as the charged precursor (8). The HO_3 that was created was shown to have a lifetime in excess of 1 μs under the collisionless conditions of these experiments, but no estimate of the $\text{HO}-\text{O}_2$ binding energy was reported. Speranza (9) examined the reactivity of HO_3^+ ions with a variety of inorganic and organic molecules using Fourier-transform ion cyclotron resonance (FT-ICR) spectrometry.

The thermochemistry of HO_3^+ was examined both from correlations between its proton-transfer efficiency and the proton affinity of selected substrates, as well as from comparisons of the electron-transfer efficiency of HO_3^+ and the standard ionization energies of the substrates. In this fashion, a value of $-4.2 \pm 21 \text{ kJ mol}^{-1}$ ($4.184 \text{ kJ mol}^{-1} \equiv 1.0 \text{ kcal mol}^{-1}$) was estimated for $\Delta_f H^\circ$ (298 K)—the standard enthalpy of formation of HO_3 —which corresponds to $D_0(\text{HO}-\text{O}_2)$ of $\sim 41 \text{ kJ mol}^{-1}$.

Lester and co-workers (10–14) have carried out a number of elegant studies of HO_3 and DO_3 using infrared action spectroscopy to examine the radicals formed in a free jet expansion. Applying energy conservation, knowledge of the energy ($h\nu$) of the photons (equivalent to $\equiv 6976.7 \text{ cm}^{-1}$) that excite the HO_3 radicals, and the energy of the highest observed OH rovibrational product channel [$\nu = 1$, $N = 8$, e (at 4840.2 cm^{-1})] enabled Murray *et al.* (10) to set an upper limit on the value $D_0(\text{HO}-\text{O}_2)$ of 25.6 kJ mol^{-1} (where ν is the frequency of the photons, ν is the vibrational quantum number of the designated level, N is the rotational quantum number, and e is a label designating the parity of the level). This value was further refined to 22.2 kJ mol^{-1} by way of

Scheme 1. Reactions involving the HO_3 radical.

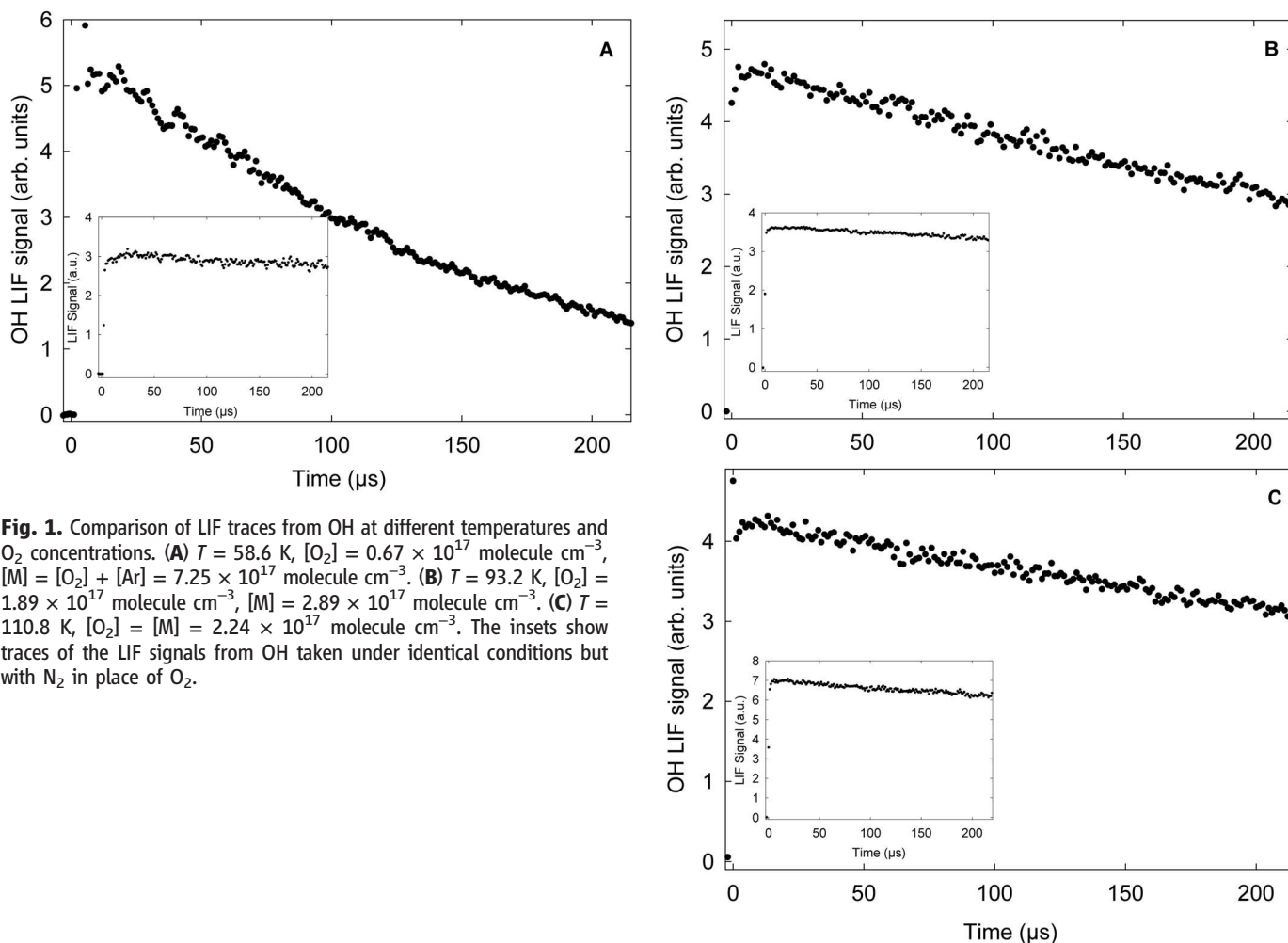
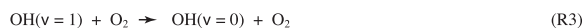
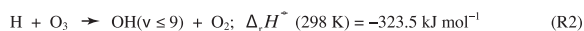


Fig. 1. Comparison of LIF traces from OH at different temperatures and O_2 concentrations. (A) $T = 58.6 \text{ K}$, $[\text{O}_2] = 0.67 \times 10^{17} \text{ molecule cm}^{-3}$, $[\text{M}] = [\text{O}_2] + [\text{Ar}] = 7.25 \times 10^{17} \text{ molecule cm}^{-3}$. (B) $T = 93.2 \text{ K}$, $[\text{O}_2] = 1.89 \times 10^{17} \text{ molecule cm}^{-3}$, $[\text{M}] = 2.89 \times 10^{17} \text{ molecule cm}^{-3}$. (C) $T = 110.8 \text{ K}$, $[\text{O}_2] = [\text{M}] = 2.24 \times 10^{17} \text{ molecule cm}^{-3}$. The insets show traces of the LIF signals from OH taken under identical conditions but with N_2 in place of O_2 .

measurements on DO_3 (14). On the basis of this later limiting value, Murray *et al.* (14) estimated values of the equilibrium constants for reaction R4 at different temperatures and hence the fraction of atmospheric OH that might be found as HO_3 at different altitudes in the Earth's atmosphere. This fraction reached a maximum value at 11 km, where it seemed that up to 25% of the OH radicals could be associated with O_2 as HO_3 radicals.

The potential importance of HO_3 in atmospheric chemistry and the difficulty of obtaining information about it from experiments have stimulated many theoretical investigations of this radical. The results of calculations performed up to 2002 were summarized by Denis *et al.* (15). Their own calculations on HOOH and HO_3 used density functional, coupled cluster, and complete basis set extrapolation methods, and through a series of calculations on the isodesmic reaction $\text{HOOH} + \text{OH} \rightarrow \text{HOH} + \text{HO}_3$ they obtained a value of $29.7 \pm 8.4 \text{ kJ mol}^{-1}$ for $\Delta_f H^\circ$ (298 K) for HO_3 . Denis and Ornellas (16) have recently extended these methods and recommend $(23.0 \pm 4.2) \text{ kJ mol}^{-1}$ for $\Delta_f H^\circ$ (298 K) for HO_3 . They also pointed out that the difference in the changes in zero-point energy for the reactions $\text{HO}_3 \rightarrow \text{OH} + \text{O}_2$ and $\text{DO}_3 \rightarrow \text{OD} + \text{O}_2$ amounts to 2.34 kJ mol^{-1} , which serves to lower the experimental upper limit for D_0 for $\text{HO}-\text{O}_2$ [derived from experiments on DO_3 (14)]

by an equivalent amount. In addition to their experiments on the rotational spectrum of HO_3 , Suma *et al.* (7) carried out multi-reference configuration interaction calculations. They obtained good agreement with the experimental rotational constants and computed the $\text{HO}-\text{O}_2$ bonding energy (D_0) to be 16.3 kJ mol^{-1} . Braams and Yu (17) have reported the results of calculations using a hybrid density functional and a large basis set aug-cc-pVTZ method and have compared their results with those of all the ab initio calculations reported up to 2008. Their calculations gave $D_e = 41.6 \text{ kJ mol}^{-1}$, a value which they claimed to be in good agreement with the experimental upper limit for D_0 measured by Murray *et al.* (10). Most recently of all, Vamer *et al.* (18) have reported the results of calculations using coupled cluster methods. Their value of D_0 is 10.5 kJ mol^{-1} . A summary of theoretical results from the compilations in references (15) and (17) is provided in table S2.

Here, we describe the results of experiments that yielded absolute values of $D_0(\text{HO}-\text{O}_2)$ and $\Delta_f H^\circ$ (HO_3). Our method is based on observations of the kinetic behavior of OH radicals, produced by means of pulsed laser photolysis (PLP) of H_2O_2 , in the presence of O_2 in the low-temperature, gas-phase environment generated in CRESU experiments (19). Relative concentrations of OH radicals are observed at different delays after their formation, using laser-induced

fluorescence (LIF) (20). At temperatures below $\sim 85 \text{ K}$, the equilibrium constant (K_c) associated with reaction R4 is sufficiently large that the concentration of OH at long times ($[\text{OH}]_\infty$) approaches a very small value; [OH] decays exponentially with time from its initial value $[\text{OH}]_0$ to essentially zero (Fig. 1A and fig. S1). Analysis of 10 such experiments between 55.7 and 79.2 K yields values of the third-order rate constants for association of OH with O_2 that increase monotonically as the temperature is lowered through this range (table S1), as would be expected for a radical association reaction proceeding over a potential energy surface with no barrier between the radical reactants and the combined species [supporting online material (SOM) text] (20).

As the temperature is raised, K_c for reaction R4 becomes smaller. The trace shown in Fig. 1C was recorded at 110.8 K—the highest temperature accessible in our experiments—with $[\text{O}_2] = 2.24 \times 10^{17} \text{ molecule cm}^{-3}$. The decay of the LIF signal from OH under these conditions is only slightly faster than that in the background experiment, in which N_2 replaces O_2 . At still higher temperatures, no removal of OH by reaction with O_2 would be observed.

In order to obtain thermodynamic data, it is necessary to examine the kinetics in the small range of intermediate temperatures at which [OH] decays to an equilibrium value significantly greater than zero—that is, to an OH concentration at long times $[\text{OH}]_\infty \sim 0.5 [\text{OH}]_0$ (Fig. 1B). Under circumstances in which the kinetics of both the forward and reverse reactions must be considered and no other processes contribute to changes in the concentrations of OH and HO_3 (21), the effective first-order rate constant (k_{1st}) associated with the exponential decay of [OH] to an equilibrium level corresponds to the sum of the pseudo-first-order rate constants for the forward and reverse reactions:

$$k_{1st} = k_{ass} [\text{O}_2] [\text{M}] + k_{diss} [\text{M}] \quad (1)$$

where [M] is the total gas density ($[\text{O}_2] + [\text{Ar}]$) and k_{ass} and k_{diss} are the third-order rate constant for association of $\text{OH} + \text{O}_2$ and the second-order

Fig. 2. Results from an experiment in which $T = 95.4 \text{ K}$, $[\text{O}_2] = 5.0 \times 10^{17} \text{ molecule cm}^{-3}$, and $[\text{M}] = 6.1 \times 10^{17} \text{ molecule cm}^{-3}$. The LIF signals from OH have been corrected for background loss by multiplying them by the factor $\exp(+k_{1st}^0 t)$. The signals are then fitted to a function of the form $\{A + B \exp(-k_{1st} t)\}$. The errors on the values of A and $A + B$ are at the 2σ level.

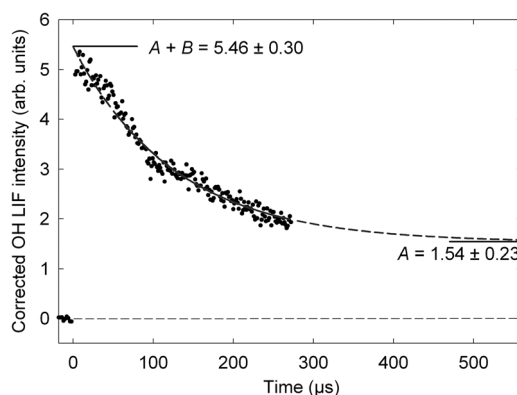


Table 1. Results from analyses of the seven experiments conducted at temperatures between 87.4 K and 99.8 K. The ratio $[\text{OH}]_\infty/[\text{OH}]_0$ is derived from the signals at zero time and long times as determined from fits to the traces of LIF signals versus time; K_c is given by Eq. 2, so $K_c [\text{O}_2] = \{1 - ([\text{OH}]_\infty/[\text{OH}]_0)/([\text{OH}]_\infty/[\text{OH}]_0)\}$;

and $Q = (1/q_{rel \text{ trans}})^{3/2} \{q_{int}(\text{HO}_3)/q_{int}(\text{O}_2) q_{int}(\text{OH})\}$. The evaluation of and errors in $\ln Q$ are discussed further in the SOM text. The errors cited in the main body of the table correspond to a single SE. The errors quoted for the values of D_0 correspond to 95% confidence limits (20).

T/K	$[\text{O}_2]/10^{17} \text{ molecule cm}^{-3}$	$k_{1st}/10^3 \text{ s}^{-1}$	$\{A/(A + B)\} = [\text{OH}]_\infty/[\text{OH}]_0$	$K_c [\text{O}_2]$	$\ln K_c$	$\ln Q$	$D_0/k \text{ mol}^{-1}$
87.4 ± 0.7	4.55 ± 0.08	12.1 ± 0.6	0.170 ± 0.008	4.90 ± 0.28	-39.07 ± 0.06	-55.376	11.85 ± 0.11
91.5 ± 0.7	4.52 ± 0.08	6.2 ± 0.7	0.147 ± 0.033	5.80 ± 1.51	-38.90 ± 0.26	-55.439	12.59 ± 0.22
94.9 ± 1.1	4.46 ± 0.14	17.1 ± 3.1	0.320 ± 0.045	2.12 ± 0.44	-39.89 ± 0.21	-55.489	12.31 ± 0.22
95.4 ± 0.7	5.00 ± 0.08	8.0 ± 0.8	0.281 ± 0.016	2.56 ± 0.21	-39.81 ± 0.08	-55.496	12.44 ± 0.11
96.7 ± 1.1	4.66 ± 0.13	13.9 ± 2.4	0.478 ± 0.033	1.09 ± 0.15	-40.59 ± 0.14	-55.514	12.00 ± 0.17
97.5 ± 0.7	4.75 ± 0.09	10.7 ± 1.9	0.464 ± 0.022	1.15 ± 0.10	-40.56 ± 0.09	-55.525	12.13 ± 0.12
99.8 ± 0.7	5.02 ± 0.09	25.6 ± 5.2	0.408 ± 0.085	1.45 ± 0.51	-40.38 ± 0.35	-55.557	12.59 ± 0.31
Weighted mean							12.18 ± 0.13
Unweighted mean							12.3 ± 0.3

rate constant for dissociation of HO₃, respectively. The equilibrium constant is

$$K_c = \frac{[\text{HO}_3]_\infty}{[\text{OH}]_\infty [\text{O}_2]} = \frac{[\text{OH}]_0 - [\text{OH}]_\infty}{[\text{OH}]_\infty [\text{O}_2]} \\ = \frac{1 - \{[\text{OH}]_\infty / [\text{OH}]_0\}}{[\text{O}_2] \{[\text{OH}]_\infty / [\text{OH}]_0\}} \quad (2)$$

Consequently, K_c can be determined from measurements of the OH concentrations, which enable estimation of the ratio $[\text{OH}]_\infty / [\text{OH}]_0$. Because the partition functions and standard molar entropies for OH, O₂, and HO₃ can be calculated by using known spectroscopic data, determination of the equilibrium constant allows one to calculate $D_0(\text{HO}-\text{O}_2)$ and $\Delta_r H^\circ$, the standard enthalpy change for reaction R4—and hence $\Delta_f H^\circ$ for HO₃—by using the standard thermodynamic Eqs. 3 and 4:

$$K_c = \left\{ \frac{Q_{\text{HO}_3}}{Q_{\text{OH}} Q_{\text{O}_2}} \right\} \exp\left(+\frac{D_0}{RT}\right) \quad (3)$$

$$K^\circ = \exp\left(\frac{\Delta_r S^\circ}{R}\right) \exp\left(\frac{-\Delta_r H^\circ}{RT}\right) \quad (4)$$

where the Q s are “per unit volume” partition functions, R is the gas constant, $\Delta_r S^\circ$ is the standard entropy change for reaction R4, and T is temperature.

This method of obtaining thermodynamic data for a radical species has been applied previously by several groups (22) to find the strength of the R–O₂ bond in alkyl peroxy radicals (RO₂). In those systems, in which the enthalpy of the R + O₂ reaction varies from ~ -75 kJ mol^{−1} (R = C₃H₅) to ~ -158 kJ mol^{−1} (R = 2-C₃H₇), the kinetics of the reaction must be studied at elevated temperatures in order to find the conditions in which $[\text{R}]_\infty \sim 0.5 [\text{R}]_0$. In the present case, in which the enthalpy change of reaction R4 is so much lower, it is necessary to perform experiments at low temperatures.

Although CRESU experiments easily achieve the necessary low temperatures, there are a number of difficulties associated with their application to the OH + O₂ reaction. These are chiefly associated with (i) the short observation time (~ 250 μs), corresponding to the limited duration of the uniform supersonic flow; (ii) the fact that R4 is a third-order reaction and there are limits to the total gas density (and hence also to the O₂ concentration) that can be achieved in a CRESU flow; and (iii) the strong quenching of the LIF signals from OH because of the high concentrations of O₂ present in our experiments.

Preliminary experiments were performed with a Laval nozzle, which when operated with pure argon generated a temperature in the supersonic flow of 52.7 K and a density of 8.2×10^{17} molecule cm^{−3}. When pure O₂ was passed through this same nozzle to create a uniform supersonic flow, the temperature was 110.8 K, and the den-

sity was 2.24×10^{17} molecule cm^{−3}. By changing the ratio of O₂ to argon in the gas, it was possible to vary the temperature in the gas jet between these two limits, although the total gas density and the concentration of O₂ were also changed. These preliminary experiments (Fig. 1) were performed at temperatures between 55.7 K and 110.8 K and established that the range of intermediate temperatures as defined above was ~ 85 to 105 K.

A Laval nozzle was then designed specifically for further experiments on the OH + O₂ reaction in this temperature range, which produced a total density in the gas flow between 4×10^{17} and 6×10^{17} molecule cm^{−3}. The temperature was tuned either by seeding the O₂ in the flow (or N₂ in control experiments) with different concentrations of argon and adjusting as necessary the pressure in the reservoir upstream from the nozzle or (over a smaller range) by varying the pressure in the main chamber. It was from experiments in which we used this nozzle that we have extracted equilibrium constants and hence the value of D_0 (20). As in the preliminary experiments (Fig. 1), we followed each measurement of the time-dependent OH LIF decay in the presence of O₂ with a corresponding control experiment in which O₂ was replaced by a flow of N₂ that generated the same temperature and density in the flow. These control experiments established the background rate at which the OH concentration decayed in the absence of O₂, a loss process that was incorporated into our analysis (20).

The LIF traces from pairs of such experiments are shown in fig. S2. Although the decays of LIF signals from OH in mixtures containing N₂ are much slower than that in the equivalent O₂/Ar mixture, they are significant. A substantial component of this background decay is likely to be due to diffusion of OH from the region illuminated by the probe laser. Although we recognize that diffusive loss is not strictly a first-order process, we believe that a suitable correction for this slow background loss can be made by including such a term, k_{loss}^0 , in the kinetic equation for the OH radical:

$$d[\text{OH}]/dt = -\{k_{\text{ass}}[\text{O}_2][\text{M}] + k_{\text{loss}}^0\}[\text{OH}] + k_{\text{diss}}[\text{M}][\text{HO}_3] \quad (5)$$

Assuming that the background loss of OH and HO₃ are described by terms with the same first-order rate constant, then it is possible (20) to derive the equation

$$[\text{OH}]_t = [\text{OH}]_0 \exp(-k_{\text{loss}}^0 t) \left\{ \frac{k_{\text{diss}}}{k_{\text{ass}}[\text{O}_2] + k_{\text{diss}}} + \frac{k_{\text{ass}}[\text{O}_2]}{k_{\text{ass}}[\text{O}_2] + k_{\text{diss}}} \exp[-k_{\text{ass}} t] \right\} \quad (6)$$

The traces of LIF signals versus time recorded in the seven pairs of experiments with the specially designed Laval nozzle were analyzed by first evaluating k_{loss}^0 from a single exponential fit to the LIF signals recorded in the experiment with N₂ and then by fitting the signals in the trace in

which O₂ was present to a function of the form $\exp(-k_{\text{loss}}^0 t) \{A + B \exp(-k_{\text{ass}} t)\}$. These fits are shown in fig. S2. The parameters A and $(A + B)$ that are derived are proportional to $[\text{OH}]_\infty$ and $[\text{OH}]_0$, which are the concentrations of OH extrapolated to long time and at zero time, respectively, in the absence of background loss. Consequently, the ratio $A/(A + B)$ corresponds to the value of $[\text{OH}]_\infty / [\text{OH}]_0$ required in Eq. 2 in order to calculate the equilibrium constant K_c . An alternative but equivalent procedure is to multiply the LIF signals by the factor $\exp(+k_{\text{loss}}^0 t)$ and fit the resultant corrected signals to the simpler function $\{A + B \exp(-k_{\text{ass}} t)\}$, as shown in Fig. 2. This plot clearly demonstrates how the fitted curve approaches a constant, nonzero value at long time.

Having obtained values of K_c , we calculated values of the bond dissociation energy (D_0) using Eq. 3 together with the partition functions for HO₃, OH, and O₂, which we estimated using spectroscopic data and standard formulas (20). Values of K_c and D_0 obtained from each of the experiments performed at temperatures between 87.4 and 99.8 K are given in Table 1. The unweighted mean of the individual values of D_0 is (12.3 ± 0.3) kJ mol^{−1}, where the cited error represents 95% confidence limits.

For some purposes, it is more useful to have a value for $\Delta_f H^\circ$ (298 K), the enthalpy of formation of HO₃, rather than D_0 . To evaluate this quantity (20), we used Eq. 3 with our value for D_0 and partition functions calculated with spectroscopic data in order to find K_c , and then K° , at 298 K. At this temperature, we found $K^\circ = (8.9 \pm 0.12) \times 10^{-4}$ (referenced to a standard state of 1 bar). Calculating $\Delta_r S^\circ$ from spectroscopic data, we then used Eq. 4 to find that the enthalpy change for reaction R4 is $\Delta_r H^\circ$ (298 K) = $-(17.9 \pm 0.3)$ kJ mol^{−1}, which, combined with $\Delta_f H^\circ$ (298 K) for OH, yields $\Delta_f H^\circ$ (298 K) = (19.3 ± 0.5) kJ mol^{−1} for HO₃. The errors in these quantities, as for that in D_0 , are given at the 95% confidence level.

The main aim of our experiments was to determine the HO–O₂ bond dissociation energy in order to use this result to estimate equilibrium constants for reaction R4 at the temperatures found in the Earth’s atmosphere and hence to estimate the fraction of OH radicals that exists as HO₃ at different altitudes. As stated earlier, Murray *et al.* (14) carried out such a calculation on the basis of a value of $D_0 = 22.2$ kJ mol^{−1}, which corresponds to the upper limit for this quantity set by their experiments. With this assumption, they found that the maximum effect on the concentration of OH would be found at ~ 11 km, where as many as 25% of the OH radicals might be associated with O₂ as weakly bound HO₃ radicals. Our value of D_0 is 9.9 kJ mol^{−1} lower than the upper limit reported by Murray *et al.* (14). The values of K_c are thus correspondingly lowered from the values reported by Murray *et al.* by a factor of $\exp(1190 \text{ K}/T)$. At 298 K with 0.2 bar of O₂ present, the equilibrium fraction of OH bound to O₂ is only $\sim 2 \times 10^{-4}$.

Repeating the calculation reported by Murray *et al.* by using the value of D_0 determined in our experiments suggests that the fraction of OH bound to O_2 is less than 0.1% at all altitudes in the Earth's atmosphere.

Consequently, the formation of HO_3 through the association of OH radicals with O_2 plays no meaningful role in the chemistry of Earth's atmosphere.

References and Notes

- U. C. Sridharan, F. S. Klein, F. Kaufman, *J. Chem. Phys.* **82**, 592 (1985).
- R. P. Wayne, *Chemistry of Atmospheres* (Oxford Univ. Press, Oxford, ed. 3, 2000).
- J. C. Polanyi, J. J. Sloan, *Int. J. Chem. Kinet.* **1**, 51 (1975).
- D. Klenerman, I. W. M. Smith, *J. Chem. Soc., Faraday Trans. II* **83**, 229 (1987).
- J. C. Polanyi, *Acc. Chem. Res.* **5**, 161 (1972).
- D. C. McCabe, B. RajaKumar, I. W. M. Smith, A. R. Ravishankara, *Chem. Phys. Lett.* **421**, 111 (2006).
- K. Suma, Y. Sumiyoshi, Y. Endo, *Science* **308**, 1885 (2005).
- F. Cacace, G. de Petris, F. Pepi, A. Troiani, *Science* **285**, 81 (1999).
- M. Speranza, *Inorg. Chem.* **35**, 6140 (1996).
- C. Murray, E. L. Derro, T. D. Sechler, M. I. Lester, *J. Phys. Chem. A* **111**, 4727 (2007).
- E. L. Derro, C. Murray, T. D. Sechler, M. I. Lester, *J. Phys. Chem. A* **111**, 11592 (2007).
- E. L. Derro, T. D. Sechler, C. Murray, M. I. Lester, *J. Chem. Phys.* **128**, 244313 (2008).
- E. L. Derro, T. D. Sechler, C. Murray, M. I. Lester, *J. Phys. Chem. A* **112**, 9269 (2008).
- C. Murray, E. L. Derro, T. D. Sechler, M. I. Lester, *Acc. Chem. Res.* **42**, 419 (2009).
- P. A. Denis, M. Kieninger, O. N. Ventura, R. E. Cachau, G. H. F. Dierksen, *Chem. Phys. Lett.* **365**, 440 (2002).
- P. A. Denis, F. R. Ornellas, *J. Phys. Chem. A* **113**, 499 (2009).
- B. J. Braams, H.-G. Yu, *Phys. Chem. Chem. Phys.* **10**, 3150 (2008).
- M. E. Varner, M. E. Harding, J. Vázquez, J. Gauss, J. F. Stanton, *J. Phys. Chem. A* **113**, 11238 (2009).
- The acronym CRESU stands for Cinétique de Réaction en Ecoulement Supersonique Uniforme, or Reaction Kinetics in Uniform Supersonic Flow. The technique was originally developed by Rowe and his co-workers for the study of ion-molecule reactions (23).
- Materials and methods are available as supporting material on Science Online.
- M. J. Pilling, P. W. Seakins, *Reaction Kinetics* (Oxford Univ. Press, Oxford, 1995) p. 195–196.
- M. J. Pilling, S. H. Robertson, P. W. Seakins, *J. Chem. Soc., Faraday Trans.* **91**, 4179 (1995) and references therein.
- B. R. Rowe, G. Dupeyrat, J. B. Marquette, P. Gaucherel, *J. Chem. Phys.* **80**, 4915 (1984).
- We are grateful to the Centre National de la Recherche Scientifique and the Université de Rennes 1 for funding.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5983/1258/DC1
Materials and Methods

SOM Text

Figs. S1 and S2

Tables S1 and S2

References

9 November 2009; accepted 19 April 2010

10.1126/science.1184459

Freshwater Outburst from Lake Superior as a Trigger for the Cold Event 9300 Years Ago

Shi-Yong Yu,^{1,2*} Steven M. Colman,^{1,3} Thomas V. Lowell,⁴ Glenn A. Milne,⁵ Timothy G. Fisher,⁶ Andy Breckenridge,⁷ Matthew Boyd,⁸ James T. Teller⁹

Paleoclimate proxy records reveal a pervasive cooling event with a Northern Hemispheric extent ~9300 years ago. Coeval changes in the oceanic circulation of the North Atlantic imply freshwater forcing. However, the source, magnitude, and routing of meltwater have remained unknown. Located in central North America, Lake Superior is a key site for regulating the outflow of glacial meltwater to the oceans. Here, we show evidence for an ~45-meter rapid lake-level fall in this basin, centered on 9300 calibrated years before the present, due to the failure of a glacial drift dam on the southeast corner of the lake. We ascribe the widespread climate anomaly ~9300 years ago to this freshwater outburst delivered to the North Atlantic Ocean through the Lake Huron–North Bay–Ottawa River–St. Lawrence River valleys.

Greenland ice-core records reveal that Earth's climate system was very unstable during the last deglaciation and early Holocene and that it experienced repeated abrupt cooling at millennial time scales (1). One of these large climate anomalies is a nearly 2°C cooling (2) in the North Atlantic sector at about 9300 calibrated years before the present (cal yr B.P.).

High-resolution and well-dated proxy records (3–7) indicate that this climate anomaly, albeit not the most prominent in the North Atlantic region (8, 9), had a Northern Hemispheric expression, with a spatial pattern nearly identical to that of the so-called 8.2-thousand-year event (10), a widespread cooling associated with the sudden drainage of the glacial Lake Agassiz–Ojibway complex through the Hudson Strait (11). This resemblance implies a common cause for the two events (12) and suggests that freshwater forcing played a prominent role in millennial-scale climate change during the last deglaciation and early Holocene (13, 14).

Numerous lines of evidence from the North Atlantic reveal substantial changes in both surface (2) and deep (15) oceanic conditions corresponding to the event at 9300 cal yr B.P. Studying the cause of this event may provide a valuable perspective on the sensitivity of the North Atlantic meridional overturning circulation (MOC) to freshwater forcing, which in turn would improve

the predictability of future climate changes due to freshwater influx associated with accelerated melting of the Arctic ice pack or the Greenland Ice Sheet. Here we present some observational constraints on the magnitude and routing of a surge of fresh water from Lake Superior at 9300 cal yr B.P.

Located in the middle of the North American continent (Fig. 1A), Lake Superior has experienced a complex evolution after the retreat of the Laurentide Ice Sheet, at about 11,500 cal yr B.P., from the southern edge of the basin, in response to glacioisostatic adjustment and outlet switching (16–18). Middle to late Holocene relative lake-level (RLL) changes in this basin are known from the dating of raised strandplain sequences (19, 20), but older RLL changes are not as well constrained (21). In contrast to this method, we reconstruct RLL changes in the Lake Superior basin by dating isolation and reconnection events recorded in sediments of small marginal lakes that were once embayments of a larger Lake Superior. Details of this method are described in the supporting online material (SOM). The isolation of a small marginal basin from the main water body could be due to changes in localized geomorphic or hydrologic conditions. For example, changes in predepositional slope can mobilize enough sediment in the coastal zone to partially limit the connectivity of a marginal basin with the major water body, thereby supporting quiet water conditions. However, the lakes used in our study are situated at substantially different elevations. The isolation of all of these elevated lakes from Lake Superior requires a RLL fall.

Continuous and undisturbed sediment sequences were recovered from frozen lake surfaces from two marginal lakes on the eastern Upper Peninsula of Michigan (Fig. 1B) and from four marginal lakes on the Sibley Peninsula of Ontario (Fig. 1C). All these basins are today hydrologically closed by a bedrock sill, except for Little Harbor, which is currently a lagoon of

¹Large Lakes Observatory, University of Minnesota Duluth, Duluth, MN 55812, USA. ²Department of Earth and Environmental Sciences, Tulane University, New Orleans, LA 70118, USA. ³Department of Geological Sciences, University of Minnesota Duluth, Duluth, MN 55812, USA. ⁴Department of Geology, University of Cincinnati, Cincinnati, OH 45221, USA. ⁵Department of Earth Sciences, University of Ottawa, Ottawa, Ontario ON K1N 6N5, Canada. ⁶Department of Environmental Sciences, University of Toledo, Toledo, OH 43606, USA. ⁷Department of Geology, Mercyhurst College, Erie, PA 16546, USA. ⁸Department of Anthropology, Lakehead University, Thunder Bay, Ontario ON P7B 5E1, Canada. ⁹Department of Geological Sciences, University of Manitoba, Winnipeg, Manitoba MB R3T 2N2, Canada.

*To whom correspondence should be addressed. E-mail: syu2@tulane.edu

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1262>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/science.1187860/DC1>

This article **cites 35 articles**, 5 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/328/5983/1262#otherarticles>

This article appears in the following **subject collections**:

Atmospheric Science

<http://www.sciencemag.org/cgi/collection/atmos>

Lake Superior with a surface elevation of 183 m above sea level (a.s.l.). In addition, a naturally exposed fluvial-lacustrine channel infill sequence in the Thunder Bay area of Ontario was studied (numbered square 7, fig. S1). More detailed information about the geographical coordinates, elevation, and radiocarbon ages of these sites is given in table S1.

Lithology, magnetic susceptibility, and loss-on-ignition data are presented in Fig. 2 along with radiocarbon ages. The latter are plotted on the right-hand side of each core. We interpret the sharp decrease in magnetic susceptibility and the corresponding increase in loss on ignition within the cores as marking the isolation of marginal basins from Lake Superior, and the reverse changes in terms of reconnection to Lake Superior. Specifically, once a marginal basin was isolated by RLL fall in Lake Superior, autochthonous and allochthonous organic matter increased in this small lake, thereby diluting the concentration of magnetic grains in the sediment delivered from the surrounding catchment, and vice versa for when the basin reconnected to Lake Superior. Thus, by knowing the elevation of these small marginal basins, along with when isolation from and/or reconnection to Lake Superior occurred,

the RLL changes in the Lake Superior basin may be constrained.

Data from Little Harbor and Lake Culhane near Whitefish Point on the eastern Upper Peninsula of Michigan (Fig. 1B) can be used to illustrate this approach. For Little Harbor, the organic-rich fibrous peat layer with extremely low magnetic susceptibilities (Fig. 2) records isolation from Lake Superior between 8365 ± 100 and 5340 ± 50 ^{14}C yr B.P. After this initial isolation, a temporary influx of clastic sediments occurred, which may be associated with a RLL rise to the middle Holocene Nipissing highstand (22). The sharp lithostratigraphic transition from silty clay to gyttja, along with a substantial rise in loss on ignition and the corresponding decrease in magnetic susceptibility at 3860 ± 35 ^{14}C yr B.P., indicates a RLL fall after the transgression, turning the embayment environment into a lagoon that otherwise would be isolated from Lake Superior because of its low elevation. Lake Culhane is 17 m higher on the landscape, and its isolation at 8425 ± 65 ^{14}C yr B.P. is recorded at 21 m depth below the lake surface.

Of the studied lakes on the Sibley Peninsula of Ontario (Fig. 1C), from highest to lowest, Ravine Lake was isolated at 8180 ± 45 ^{14}C yr

B.P. and Grassy Lake at 8210 ± 50 ^{14}C yr B.P. For Roadside Lake, we are unable to determine when the initial isolation occurred during the early Holocene, because the older sediments were scoured away in this channel-like basin while RLL was falling. However, the results show the lake reconnecting to Lake Superior during the RLL rise to the Nipissing highstand and being isolated again at 3600 ± 40 ^{14}C yr B.P. A thin gyttja layer sandwiched in the thick silty clay may indicate a short-lived inwashing event. Surprise Lake was initially isolated at 8170 ± 40 ^{14}C yr B.P., reconnected to Lake Superior at 8010 ± 40 ^{14}C yr B.P., and was finally isolated at 1900 ± 65 ^{14}C yr B.P. The RLL fall from the Nipissing highstand may explain the final isolation. At Old Fort William, in an embayment of Lake Superior near Thunder Bay, a substantial drawdown of the lake is recorded by channel incision into fluvial sediments of the Kaministiquia delta. Radiocarbon ages of terrestrial macrofossils preserved in the bottom of this paleochannel (at 187 m a.s.l.) indicate that a low water level was established there before 8000 ^{14}C yr B.P. (Fig. 2).

Taken together, our results indicate a general trend of RLL fall on both the eastern Upper

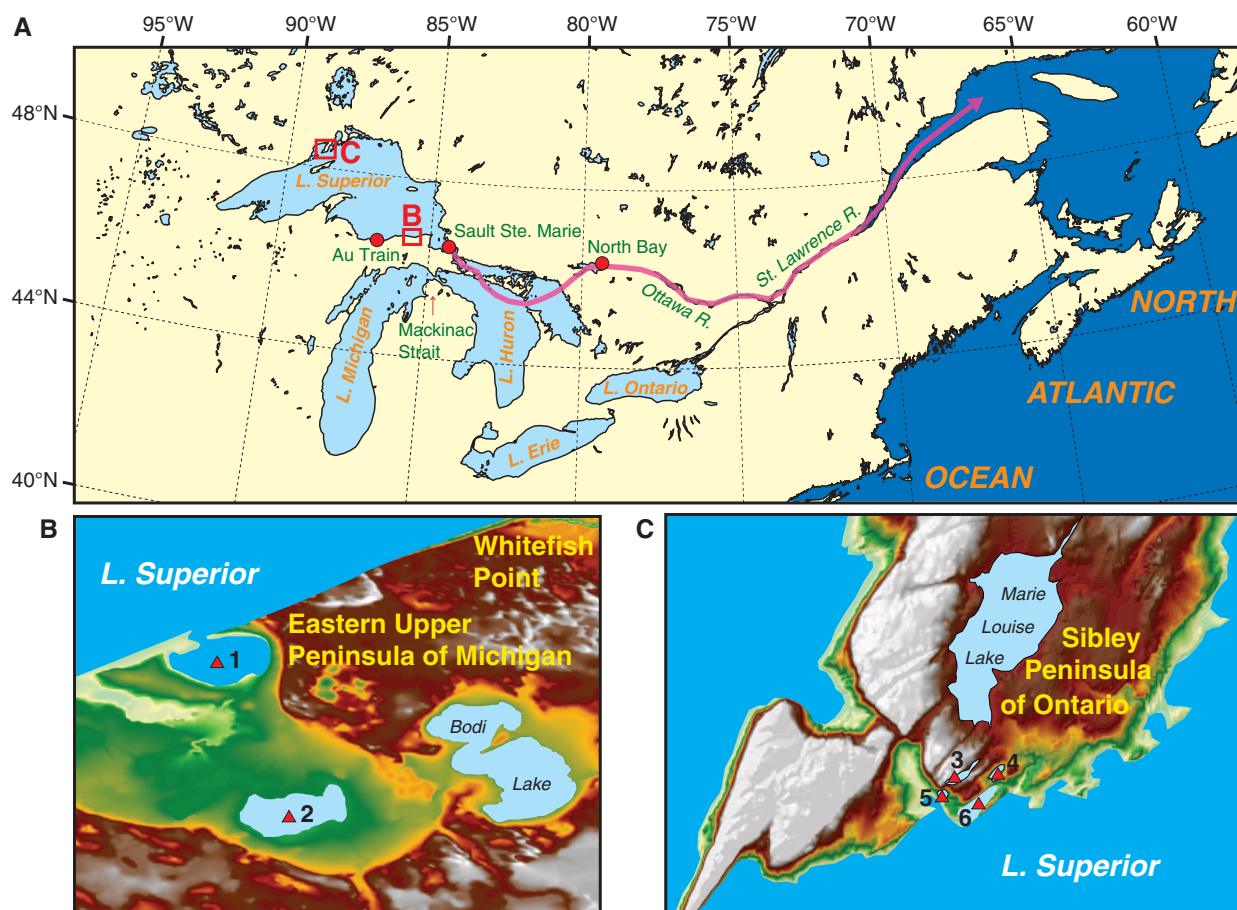


Fig. 1. (A) Location of the Laurentide Great Lakes system and its connecting river channels. Red solid dots indicate outlets mentioned in the text. The arrow denotes the proposed route of the Lake Superior flood through the Lake Huron–North Bay–Ottawa River–St. Lawrence River valleys that was coeval

with the event at 9300 cal yr B.P. (B) Shaded relief map showing the present-day topography of the eastern Upper Peninsula of Michigan. 1, Little Harbor; 2, Lake Culhane. (C) Modern shaded relief map of the Sibley Peninsula of Ontario. 3, Lake Ravine; 4, Lake Grassy; 5, Lake Roadside; 6, Lake Surprise.

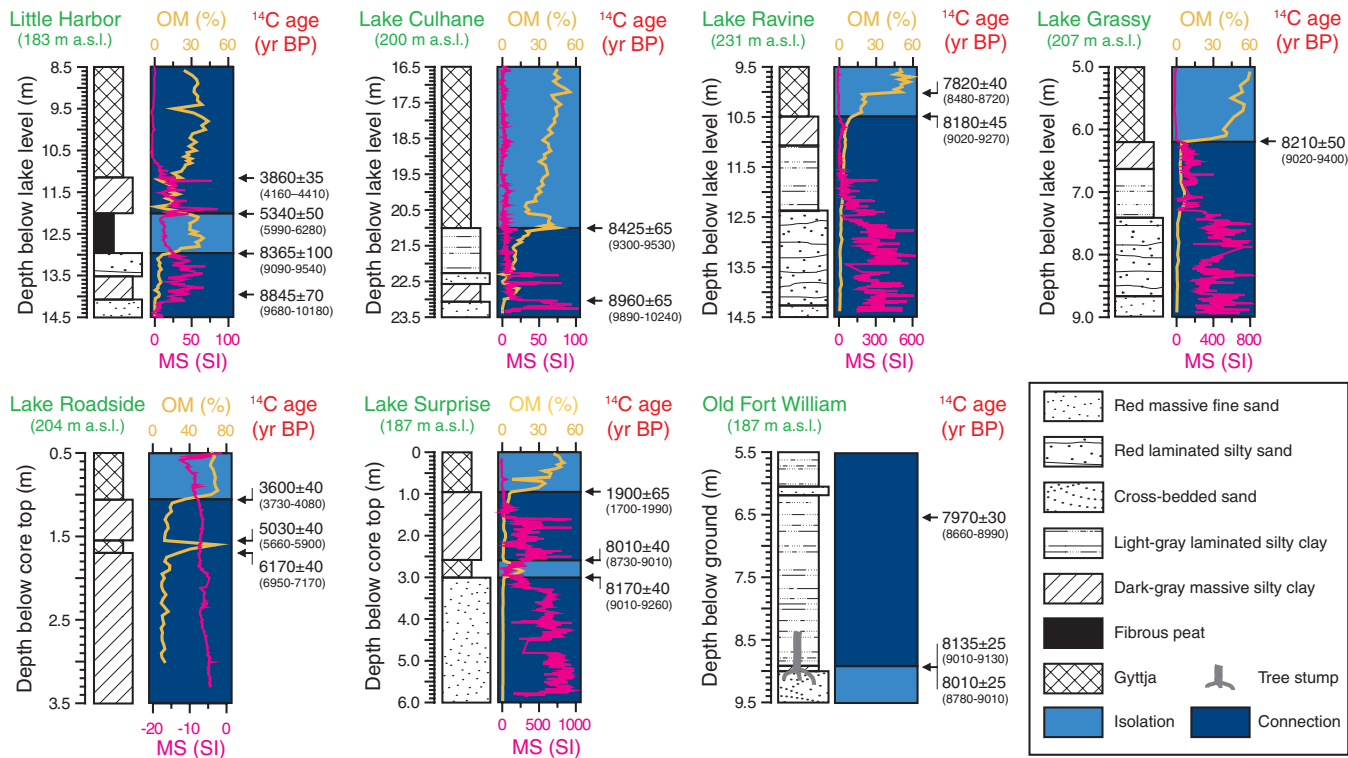


Fig. 2. Lithostratigraphic logs of lacustrine sediment cores from the eastern Upper Peninsula of Michigan and Sibley Peninsula of Ontario, as well as a fluvial-lacustrine channel infill sequence from the Old Fort William cutbank site, Ontario. Arrows indicate accelerator mass spectrometry ¹⁴C-dated levels. Detailed in-

formation about the cores and radiocarbon ages is in table S1. The identification of lake status in terms of isolation and connection with Lake Superior is based on magnetic susceptibility (MS) and organic matter (OM) measurements. Numbers in brackets are calibrated radiocarbon ages with 2σ error.

Peninsula of Michigan and Sibley Peninsula of Ontario until about 9000 cal yr B.P., followed by a RLL rise thereafter. To get a broader perspective on long-term RLL changes, we compiled published data from around Lake Superior for ¹⁴C-dated lake isolation and reconnection events as well as for fluvial/lacustrine sediment contacts in estuaries (fig. S1 and table S2). With detailed information on these RLL indicators, we reconstruct RLL changes in the Lake Superior basin by synthesizing these data with our results.

To estimate the magnitude of RLL changes from these data sites and place them into a time-varying rate-of-change context, each site must be corrected for differential uplift with respect to a reference site (SOM). We predicted land uplift at each site during the past 11,500 years due to glacioisostatic adjustment, using the ICE-3G ice loading model (23) and a multilayered, radially symmetric, and self-gravitating Maxwell Earth model (24). We also used the ICE-5G ice model (25), but predictions based on it were less compatible with the instrumental data (26) than were those generated using ICE-3G (fig. S2). The elastic and density structure of the Earth model is based on seismic constraints (27), and the viscous structure was varied to explore the sensitivity of the predictions to this aspect of the model. The algorithm used to compute vertical land motion is based on that described in Mitrovica *et al.*

Fig. 3. Holocene RLL change at Sault Ste. Marie, southeastern Lake Superior, calculated by correcting observed paleo-RLL heights using an iso-static Earth model. Horizontal error bars are defined by tree-ring calibration of ¹⁴C ages with 2σ error. Vertical bars denote maximum error as a sum of uncertainties arising from both elevation estimates and the indicative range of RLL indicators. Elevation of the Nadoway barrier is estimated from the lowest point of the present-day surface (271 m a.s.l.). The difference in elevation between this barrier and the bedrock sill at Sault Ste. Marie (181 m a.s.l.) provides a maximum-limiting constraint on the rapid RLL fall about 9300 years ago.

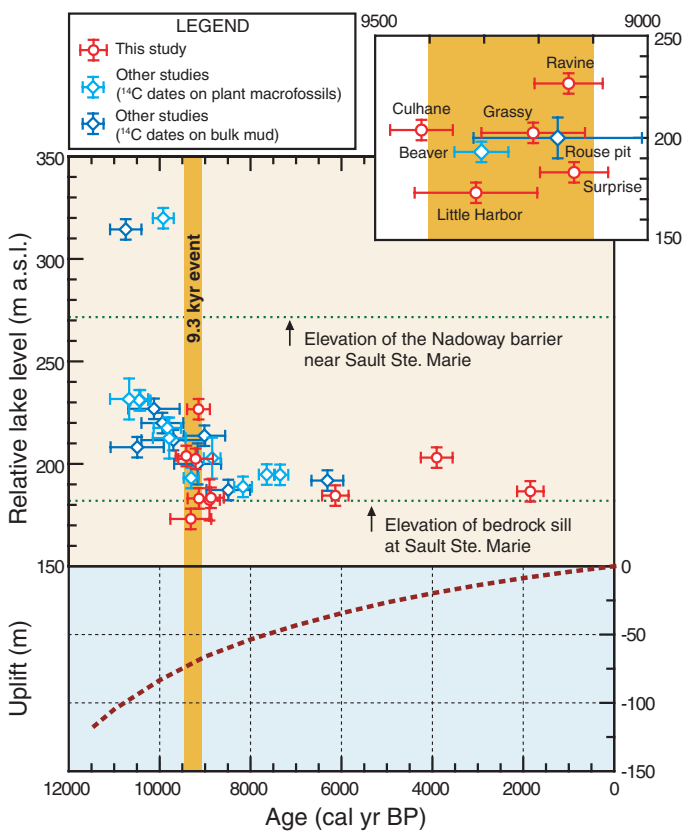
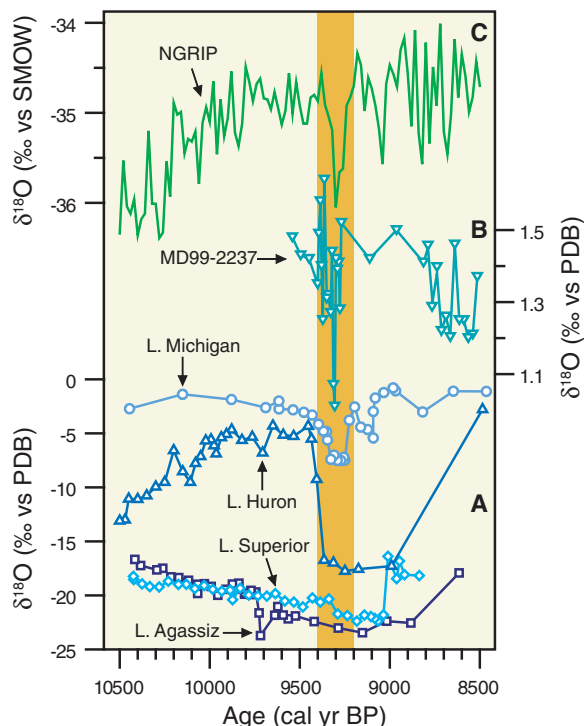


Fig. 4. Comparison of $\delta^{18}\text{O}$ records from Lake Agassiz (32), the upper Great Lakes (18, 31, 33), core MD99-2237 from the southwestern Labrador Sea (36, 37), and an ice core from the North Greenland Ice Core Project (NGRIP) (1). The vertical bar denotes a substantial depletion of $\delta^{18}\text{O}$ corresponding to the event at 9300 cal yr B.P. PDB, Pee Dee belemnite standard; SMOW, standard mean ocean water.



(28, 29), but extended to include the influence of Earth rotation (30). Varying the viscosity structure substantially affects the amplitude and rate of land uplift at specific sites (fig. S3); however, predictions of uplift between sites are remarkably insensitive to a wide range of plausible viscosity structures (fig. S4). This is a key point because it means that our corrections for differential uplift are largely independent of this poorly defined model parameter.

Sault Ste. Marie (Fig. 1A) was arbitrarily chosen as a reference site for our basin-scale RLL reconstruction. Today, it functions as the outlet draining Lake Superior to the North Atlantic Ocean through the lower Great Lakes and the St. Lawrence River valley. The modeled RLL projected to Sault Ste. Marie fell progressively from about 11,000 to 9300 cal yr B.P. (Fig. 3), primarily controlled by the isostatic rebound of the deformed basins. Overflow from the Lake Superior basin may have flowed into Lake Michigan through the Au Train–Whitefish channel before 10,600 cal yr B.P. (31). Subsequent isostatic rebound may have closed this outlet, and Lake Superior was dammed up by a drift moraine across the Nadoway barrier near Sault Ste. Marie (21), keeping the RLL higher at this time than the low-water Chippewa and Stanley phases in the Lake Michigan and Huron basins, respectively (17). Although earlier studies have suggested that the outlet across the Nadoway dam was progressively lowered by erosion (22), we suggest that the drift moraine dam was eroded and breached abruptly, as suggested by our lake-isolation data, which indicate an ~45-m rapid RLL fall from 226 to 181 m a.s.l. relative to Sault Ste. Marie between 9400 and 9100 cal yr

B.P. (inset, Fig. 3). The rapidity and magnitude of this RLL fall are independent of the choice of reference site used for the corrections of differential uplift (fig. S5).

The $\delta^{18}\text{O}$ records from Lake Agassiz (32) and the upper Great Lakes (18, 31, 33) support our interpretation of a sudden RLL drop about 9300 years ago. We ascribe the large and abrupt negative shift of the Lake Huron $\delta^{18}\text{O}$ record at ~9300 cal yr B.P. (Fig. 4A) to this ~45-m rapid RLL fall, due most likely to a sudden failure of the Nadoway drift moraine dam, which would have released about $4 \times 10^{12} \text{ m}^3$ of isotopically light glacial meltwater into Lake Huron through the St. Marys River. Increased mass exchange between Lakes Huron and Michigan may have occurred through the Mackinac Strait, as indicated by a substantial decrease of $\delta^{18}\text{O}$ values in the Lake Michigan record (31) (Fig. 4A). Although our chronology of marginal lakes in the Lake Superior basin and the isotope chronology in several of the Great Lakes' basins constrain the duration of this event only within a century or two, the breach of a sandy moraine dam may have occurred rapidly, and if within 1 year, a RLL fall of 45 m is equivalent to a flux of ~0.15 sverdrup (1 sverdrup = $10^6 \text{ m}^3/\text{s}$). Because there are no substantial changes in the $\delta^{18}\text{O}$ records from Lakes Erie (34) and Ontario (35) during this period, we infer that this cataclysmic outflow was delivered through the Ottawa River–St. Lawrence River route to the North Atlantic Ocean (Fig. 1A). Although previous to this outburst flood, water from Lake Superior was also being routed to the North Bay–Ottawa–St. Lawrence River; the sudden decrease of $\delta^{18}\text{O}$ in the surface waters of the southwestern Labrador Sea at about

9300 cal yr B.P. (36, 37) can be explained by a short-lived flood (Fig. 4B). Within dating uncertainties, we propose that it was the rapid drawdown of water in the Lake Superior basin that caused the widespread abrupt cooling in the North Atlantic (Fig. 4C) and elsewhere in the Northern Hemisphere (3–7).

We provide evidence for a rapid and large drawdown of Lake Superior, due most likely to the sudden failure of an end-moraine dam at its southeastern corner at about 9300 cal yr B.P. The catastrophic breaching of this glacial drift dam lowered the RLL by about 45 m, releasing a surge of fresh water to the North Atlantic Ocean, thereby leading to a widespread cooling. Our findings not only support the hypothesis of freshwater forcing of rapid climate changes during the last deglaciation and early Holocene, but also provide quantitative control on the sensitivity of the North Atlantic MOC to freshwater perturbation within the context of a relatively warm early Holocene climate.

References and Notes

1. S. O. Rasmussen, B. M. Vinther, H. B. Clausen, K. K. Andersen, *Quat. Sci. Rev.* **26**, 1907 (2007).
2. R. E. Came, D. W. Oppo, J. F. McManus, *Geology* **35**, 315 (2007).
3. G. H. Haug, K. A. Hughen, D. M. Sigman, L. C. Peterson, U. Röhl, *Science* **293**, 1304 (2001).
4. C. A. Dykoski et al., *Earth Planet. Sci. Lett.* **233**, 71 (2005).
5. D. Fleitmann et al., *Quat. Sci. Rev.* **26**, 170 (2007).
6. B. Cai et al., *Geophys. Res. Lett.* **35**, L02705 (2008).
7. R. Boch, C. Spötl, J. Kramers, *Quat. Sci. Rev.* **28**, 2527 (2009).
8. J. D. Marshall et al., *Geology* **35**, 639 (2007).
9. U. von Grafenstein, H. Erlenkeuser, A. Brauer, J. Jouzel, S. J. Johnsen, *Science* **284**, 1654 (1999).
10. R. B. Alley et al., *Geology* **25**, 483 (1997).
11. D. C. Barber et al., *Nature* **400**, 344 (1999).
12. D. Fleitmann et al., *Paleoceanography* **23**, PA1102 (2008).
13. P. U. Clark et al., *Science* **293**, 283 (2001).
14. J. R. Thornalley, I. N. McCave, H. Elderfield, *Paleoceanography* **25**, PA1201 (2010).
15. D. W. Oppo, J. F. McManus, J. L. Cullen, *Nature* **422**, 277 (2003).
16. C. F. M. Lewis, T. W. Anderson, *J. Paleolimnol.* **2**, 99 (1989).
17. C. F. M. Lewis, S. M. Blasco, P. L. Gareau, *Geogr. Physique Quaternaire* **59**, 187 (2005).
18. A. Breckenridge, T. C. Johnson, *Quat. Res.* **71**, 397 (2009).
19. C. E. Larsen, *J. Great Lakes Res.* **20**, 108 (1994).
20. J. W. Johnston, S. J. Baedke, R. K. Booth, T. A. Thompson, D. A. Wilcox, *J. Great Lakes Res.* **30**, 1 (2004).
21. M. Saarnisto, *Can. J. Earth Sci.* **12**, 300 (1975).
22. W. R. Farrand, C. W. Drexler, Eds., *Late Wisconsinan and Holocene History of the Lake Superior Basin* (Special Paper 30, Geological Association of Canada, St. John's, Canada, 1985), pp. 17–32.
23. A. M. Tushingham, W. R. Peltier, *J. Geophys. Res.* **96**, 4497 (1991).
24. W. R. Peltier, *Rev. Geophys.* **12**, 649 (1974).
25. W. R. Peltier, *Annu. Rev. Earth Planet. Sci.* **32**, 111 (2004).
26. A. Mainville, M. R. Craymer, *Geol. Soc. Am. Bull.* **117**, 1070 (2005).
27. A. M. Dziewonski, D. L. Anderson, *Phys. Earth Planet. Inter.* **25**, 297 (1981).
28. J. X. Mitrovica, J. L. Davis, I. I. Shapiro, *J. Geophys. Res. Solid Earth* **99**, 7057 (1994).
29. J. X. Mitrovica, J. L. Davis, I. I. Shapiro, *J. Geophys. Res. Solid Earth* **99**, 7075 (1994).

30. J. X. Mitrovica, G. A. Milne, J. L. Davis, *Geophys. J. Int.* **147**, 562 (2001).
31. S. M. Colman, L. D. Keigwin, R. M. Forester, *Geology* **22**, 547 (1994).
32. T. C. Moore Jr. *et al.*, *Paleoceanography* **15**, 4 (2000).
33. D. K. Rea *et al.*, *Can. J. Earth Sci.* **31**, 1586 (1994).
34. M. J. S. Tevesz, A. L. Sponberg, J. A. Fuller, *J. Paleolimnol.* **20**, 295 (1998).
35. M. A. McFadden, H. T. Mullins, W. P. Patterson, W. T. Anderson, *Limnol. Oceanogr.* **49**, 1570 (2004).
36. C. Hillaire-Marcel, A. de Vernal, D. J. W. Piper, *Geophys. Res. Lett.* **34**, L15601 (2007).
37. A. E. Carlson *et al.*, *Nat. Geosci.* **1**, 620 (2008).
38. S.Y.Y. thanks the University of Minnesota Duluth for providing financial support for this work. T.V.L. thanks the Comer Science and Education Foundation for support of this work. G.A.M. acknowledges support from the Canada Research Chairs program and the Natural Sciences and Engineering Research Council (NSERC) of Canada. J.T.T. thanks the NSERC of Canada for support through the Discovery Grants Program. We are indebted to the Sleeping Giant National Park, Ontario, for allowing us access to and coring of the lakes on southern Sibley Peninsula of Ontario. Our gratitude is also extended to A. Noren, A. Myrbo, and K. Brady at the National Lacustrine Core Repository, University of

Minnesota Twin Cities, for their kind assistance in core logging, and to two anonymous reviewers for comments.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1187860/DC1
Materials and Methods

Figs. S1 to S5

Tables S1 and S2

References

2 February 2010; accepted 23 April 2010

Published online 29 April 2010;

10.1126/science.1187860

Include this information when citing this paper.

Fractal Organic Hazes Provided an Ultraviolet Shield for Early Earth

E. T. Wolf and O. B. Toon

The Archean Earth (3.8 to 2.5 billion years ago) was probably enshrouded by a photochemical haze composed of fractal aggregate hydrocarbon aerosols. The fractal structure of the aerosols would have had a strong effect on the radiative properties of the haze. In this study, a fractal aggregate haze was found to be optically thick in the ultraviolet wavelengths while remaining relatively transparent in the mid-visible wavelengths. At an annual production rate of 10^{14} grams per year and an average monomer radius of 50 nanometers, the haze would have provided a strong shield against ultraviolet light while causing only minimal antighreenhouse cooling.

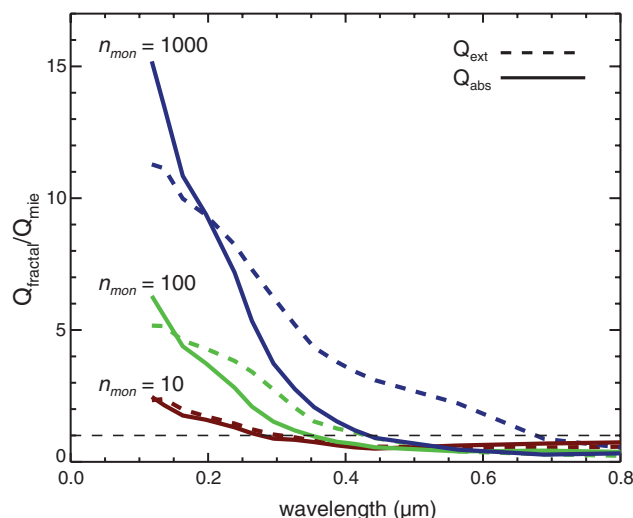
The atmosphere of the Archean Earth (3.8 to 2.5 billion years ago) was much different from the present one. The prevailing view of the Archean atmosphere is that it consisted primarily of N_2 with lesser amounts of CO_2 , CH_4 , H_2 , and H_2O . Solar evolutionary models predict that the young Sun was up to ~30% less luminous than now (1), but the lack of evidence of glaciation combined with positive evidence of primitive life indicates that Archean surface temperatures were generally as warm or warmer than today (2, 3). Resolving this paradox is important for understanding the environment of Earth at the time of the origin of life. Although a dense CO_2 atmosphere could theoretically have warmed the early Earth, the absence of siderite in fossil weathering profiles constrains the amount of CO_2 present in the young atmosphere (4). Combined greenhouse warming from CO_2 , CH_4 , NH_3 , and other less abundant gases is typically invoked to resolve this apparent paradox. Past studies indicate that N_2 - CH_4 photochemistry produces an organic haze at high altitudes that could cool the planet, offsetting any greenhouse warming (5, 6). In contrast, we show here that the haze would be optically thin at visible wavelengths and therefore have little cooling effect on climate, but would be optically thick at ultraviolet (UV) wavelengths and thus could shield reduced gases from photolysis. The key is to consider the fractal nature of the haze particles.

Atmospheric conditions similar to those that prevailed on the young Earth exist currently on Titan, a moon of Saturn. Titan's atmosphere is strongly reducing, consisting of 98.4% N_2 along with 1.6% CH_4 and other hydrocarbons. Photochemical production of organic aerosols on Titan produces an optically thick haze that has been much studied (7). Similar organic aerosols are observed to form readily even in weakly reducing early-Earth-like laboratory gas mixtures having CO_2 : CH_4 ratios as high as 10:1 (8). A Titan-like photochemical haze should have formed high in the Archean atmosphere as well. Sagan and

Chyba (9) argued that an organic haze may have provided a strong UV shield, protecting underlying constituents and primitive organisms from photochemical destruction. It is of particular interest that if NH_3 is protected by a strongly UV-absorbing haze and allowed to accumulate, a surface mixing ratio of 10^{-5} could alone provide sufficient greenhouse warming to keep Archean surface temperatures above freezing (1). However, recent modeling studies have suggested that the haze would be of a similar optical thickness in both the UV and visible wavelengths (5, 6). Thus, any increase in UV shielding is accompanied by a thickening at visible wavelengths, which would cause intense antighreenhouse cooling and freeze the planet.

Previous models of the Archean haze have inaccurately made the simplifying assumption of particle sphericity. Hydrocarbon aerosols have long been known to form fluffy aggregates that exhibit fractal structure (10). This fractal structure affects both the microphysical and radiative properties of the haze. Given the expected chemical and microphysical similarities between Titan and early Earth, we simulated haze particles as fractal aggregates (11), using a scheme introduced by Cabane *et al.* (12) and later used by Rannou *et al.* (13–15), to model fractal aggregate hazes in Titan's atmosphere. The fractal model is successful in reproducing the scattering properties of Titan's haze (7).

Fig. 1. The ratio of absorption (Q_{abs} , solid line) and extinction (Q_{ext} , dashed line) efficiencies calculated for fractal aggregate particles ($Q_{fractal}$) to those calculated for Mie particles of equal mass (Q_{mie}). A value of unity indicates that the aggregate particle is radiatively indistinguishable from the Mie particle. The particles illustrated have $D_f = 2$, $r_{mon} = 50$ nm, and $n_{mon} = 10$, 100, and 1000, corresponding to fractal radii of 0.16, 0.5 and 1.6 μm , respectively (Eq. 1). We assume refractive indices for Titan analogs (19). Fractal aggregate particles are an order of magnitude more absorbing in the UV while being slightly less absorbing in the visible than are Mie particles of equal mass. The gradient increases for larger aggregates containing more monomers.



Department of Atmospheric and Oceanic Sciences, Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO 80309-0311, USA.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1266>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1266/DC1>

This article **cites 23 articles**, 4 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/328/5983/1266#otherarticles>

This article has been **cited by** 1 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/328/5983/1266#otherarticles>

This article appears in the following **subject collections**:

Atmospheric Science

<http://www.sciencemag.org/cgi/collection/atmos>

A fractal aggregate is formed by the coagulation of many primary spherical particles (henceforth called monomers) into a larger complex structure. Although the shape of each individual aggregate can appear complicated, fractal objects have the unifying property of self-similarity. The structural properties of fractal aggregates can be quantified to a first approximation by a single parameter, the fractal dimension, denoted D_f . The fractal dimension describes the morphological dimension of the aggregate: $D_f = 3$ describes a compact spherical particle, whereas $D_f = 1$ describes a linear chain. When $D_f < 3$, the particle radius is not so easily defined. Fractal geometry provides the needed mathematical framework to proceed. The radius of a fractal aggregate, R_f , is defined by the equation

$$n_{\text{mon}} = \alpha \left(\frac{R_f}{r_{\text{mon}}} \right)^{D_f} \quad (1)$$

where n_{mon} is the number of monomers contained in the aggregate, α is a dimensionless constant

(taken here to be unity), and r_{mon} is the average monomer radius (16).

Fractal particles behave optically like a bimodal distribution of spherical particles. Monomers contained within the aggregate interact strongly with shortwave radiation, whereas the bulk size of the aggregate affects longer wavelengths. Considerable work has been done on the theory of scattering and absorption by fractal aggregate particles. Here we adopt the mean-field approximation of Mie scattering by fractal aggregates of identical spheres proposed by Botet *et al.* (17, 18). In this method, fractal aggregates are considered to be composed of n_{mon} identical spherical monomers of radius r_{mon} , each acting as a Mie particle. The total scattered electric field from an aggregate is the sum of the scattered fields from each monomer contained within the aggregate. We assume refractive indices for Titan-analog hazes (19). The incorporation of oxygenated species into early-Earth hazes would probably have altered their refractive indices; however, wavelength-dependent data for oxygenated hazes are not currently available. As illustrated in Fig. 1, fractal

aggregates are considerably more absorbing in the UV while being slightly more transparent in the visible and near-infrared than are Mie particles of equal mass. This trend becomes more pronounced as n_{mon} increases and is attributed to the inherent bimodal nature of fractal aggregates (20).

In this study, we focused on an annual haze production rate of $10^{14} \text{ g year}^{-1}$ because it presents the most interesting ramifications for the Archean climate and is also a median value predicted for the production rate of early-Earth hazes (8). For comparison, the current rate of mass production in the global S cycle is about $1 \times 10^{14} \text{ g year}^{-1}$ to $2 \times 10^{14} \text{ g year}^{-1}$ (21), whereas the current rates of organic C burial and of CH_4 production are of the same order of magnitude (5). Measurements of the phase function and polarization of Titan fractal aggregate hazes constrain monomer radii to be no larger than 50 nm (7). We used the canonical value of $r_{\text{mon}} = 50$. Simulations were conducted using both the fractal and spherical models.

Using a size-resolved aerosol model in a three-dimensional global climate model (11), we are able to predict the global distribution of the particle optical and microphysical properties. Particle effective radii reach micrometer sizes in the lower atmosphere in these simulations; such sizes are similar to those that occur in the atmosphere of Titan (Fig. 2). The effective radii of equal-mass spheres, R_b , are an indicator of the mass per aggregate rather than its true geometric size. In the troposphere, R_b is larger than the effective radii of spherical particles, R_s , by a factor of 1.5, indicating that fractal aggregates grow more massive than do spherical particles. This trend is attributed to the increased geometric cross section of aggregates, which both reduces sedimentation velocities and enhances coagulation. The effective fractal radius, R_f , describes geometrical information about fractal aggregates. In the troposphere, $R_f > R_s$ by a factor of 4 and is more than double R_b . Our results indicate that fractal aggregates grow to significantly greater geometric sizes while remaining only slightly more massive than spherical particles, thus reflecting the fluffy nature of fractal aggregates. We find that early-Earth haze particles typically would have contained $\sim 10^4$ monomers (11). For comparison, Titan aerosols are estimated to contain ~ 3000 monomers per aggregate in its stratosphere; however, estimates are not well constrained because of uncertainties in the fractal dimension, average monomer size, and haze optical constants, all of which are required to determine n_{mon} from observations (7).

The global mean effective optical depth (τ_{eff}) was calculated using both spherical and fractal models (22). The spherical simulation predicts that the early-Earth haze would have been too thin in the UV ($\tau_{\text{UV}} = 0.73$) to provide effective shielding, but sufficiently thick in the visible ($\tau_{\text{vis}} = 0.55$) to initiate mild antigreenhouse cooling. The ratio of the effective optical depth in the UV to

Fig. 2. The globally and annually averaged effective radii for spherical particles (solid line), equal-mass sphere effective radii calculated for aggregate particles (dashed line), and the effective fractal radii (dash-dotted line) are shown for a haze production rate of $10^{14} \text{ g year}^{-1}$ and $r_{\text{mon}} = 50$ nm. Fractal aggregates grow both more massive and geometrically larger than spherical particles.

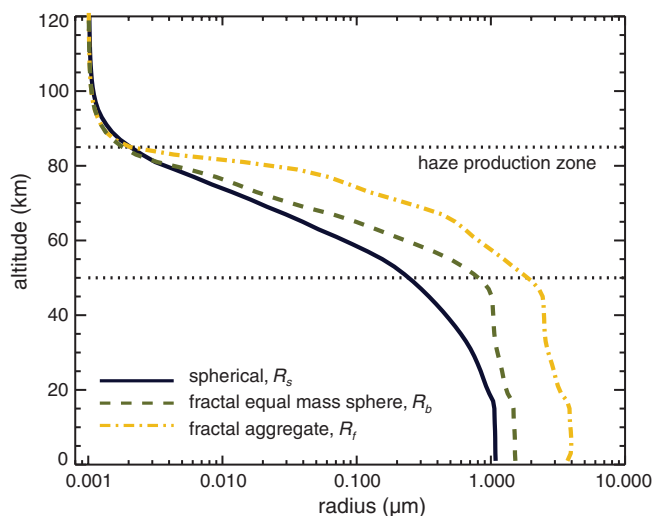
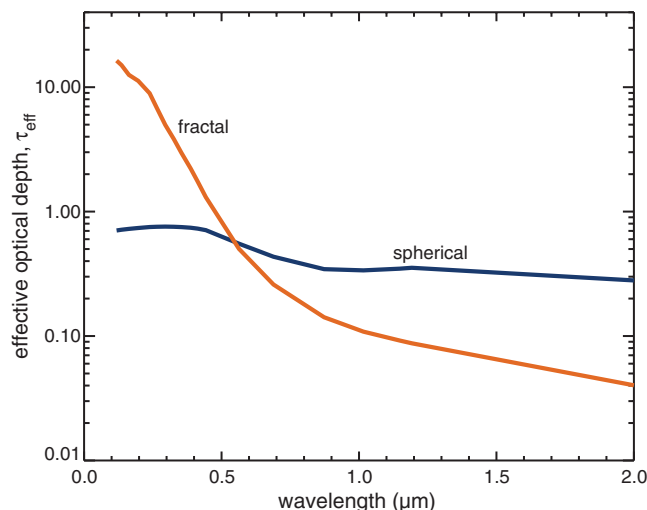


Fig. 3. The globally and annually averaged effective optical depth (τ_{eff}) plotted for both spherical and fractal models for a haze production rate of $10^{14} \text{ g year}^{-1}$ and $r_{\text{mon}} = 50$. Fractal aggregate particles (orange line) provide strong UV shielding while remaining thin at visible wavelengths. A steep gradient is exhibited at wavelengths below $0.8 \mu\text{m}$. A spherical haze (blue line) exhibits no such gradient, shielding all wavelengths more or less equally.



that in the visible is $\tau_{UV}/\tau_{vis} \approx 1.3$. In this case, the young Earth would be cooled by the haze while not benefiting from a substantial accumulation of photolytically shielded greenhouse gases. This result agrees with previous studies that assumed particle sphericity (5, 6).

The fractal simulation indicates that the haze could have been a highly effective UV shield. Organic aerosols produced at a rate of 10^{14} g year⁻¹ would have sustained a haze that was optically thick in the UV while remaining optically thin in the mid-visible (Fig. 3). For $r_{mon} = 50$ nm, $\tau_{UV} = 11.23$ and $\tau_{vis} = 0.50$, thus the ratio of effective optical depths is $\tau_{UV}/\tau_{vis} \approx 22.4$. A fractal aggregate haze is optically thinner in the mid-visible and infrared while being more than an order of magnitude thicker in the UV when compared to the spherical particle model. Sensitivity tests indicate that τ_{UV}/τ_{vis} remains large for all feasible haze production rates; however, a haze production rate near 10^{14} g year⁻¹ is required to maintain a haze that is both sufficiently thick in the UV to protect reduced gases from photolysis and sufficiently thin in the visible as to not cause substantial antigreenhouse cooling (11). Additional sensitivity tests indicate that our results are largely independent of the choice of r_{mon} , D_f , and optical constants (11). This work marks an important result in the study of the Archean climate, because it suggests a viable means of creating a UV shield that would allow the buildup of reduced gases without causing strong antigreenhouse cooling.

The mass independent fractionation of S isotopes found in Archean sediments is cited as evidence against the presence of a high-altitude UV shield (23). We do not believe this poses a serious contradiction, because a fractal aggregate haze will shield only the lower atmosphere (altitudes less than 20 km). Photolysis of SO₂ above the haze may produce the observed S isotopic signatures.

This work revitalizes the Miller-Urey hypothesis that the early Earth was home to a strongly reducing atmosphere (24). In particular, the presence of NH₃ is of dual importance. As previously mentioned, NH₃ mixing ratios of 10^{-5} or more added to the Archean atmosphere could alone resolve the faint young Sun paradox (1). However, without UV shielding, an ammonia mixing ratio of 10^{-5} would be irreversibly converted into N₂ in less than 10 years (25). Ammonia resupply rates during the Archean could not have supported an ammonia mixing ratio above 10^{-8} without a strong UV shield (26).

Sagan and Chyba estimated the atmospheric lifetime of NH₃ by

$$t_{NH_3} = \frac{t'_{NH_3}}{e^{-\tau_{UV, sec \theta}}} \quad (2)$$

where t_{NH_3} is the atmospheric lifetime of NH₃ in the presence of an overlying haze, t'_{NH_3} is the atmospheric lifetime of NH₃ in the absence of haze, τ_{UV} is the haze optical depth at 200 nm, and

θ is the solar zenith angle, taken here to be 45° (9). In this work, the haze effective optical depth at 200 nm is found to be ~ 11 . In this case for $[NH_3] \sim 10^{-5}$, the atmospheric lifetime is extended to 7×10^7 years; thus, ammonia could feasibly accumulate to high concentrations and persist throughout much of the Archean period even if sources were small. Granted, this number is probably an overestimation of the true lifetime of ammonia, because the calculation assumes that all NH₃ lies beneath the UV shield. One-dimensional photochemical modeling indicates that for $[NH_3] \sim 10^{-5}$, the altitude of peak photolysis occurs near 30 km (25), whereas in this study the haze is most absorbing near 20 km. The altitude of peak photolysis is critically dependent on the chosen eddy diffusion profile (27). Past models have assumed contemporary values for the eddy diffusion coefficient; however, a strongly absorbing haze may stabilize the upper troposphere/lower stratosphere, reducing the vertical transport of constituents to above the haze. Ammonia may have additional sinks, such as removal in rainfall, that we do not consider here but may ultimately control its lifetime.

Ammonia plays an important role in prebiotic synthesis. Reducing atmospheres produce amino acids and other organic molecules in spark discharge experiments (24). More-oxidizing atmospheres are not conducive to prebiotic synthesis of complex organics. In the view of this work, the strong UV shielding characteristics of a fractal aggregate haze would allow the young atmosphere to remain reducing while not causing substantial antigreenhouse cooling, thus making the atmosphere/ocean system a favorable environment for prebiotic synthesis. The degree to which the Archean atmosphere was reducing is dependent on the redox state of the young mantle, which is not well constrained but is believed to have been more reduced than today (28). The haze itself would be composed of complex organic molecules that would precipitate down into the primordial oceans, providing a source of organics to the surface comparable to current carbon burial rates. A strongly UV-shielding haze is also desirable because it would protect young organisms from the intense high-frequency radiation of the young Sun. A fractal aggregate haze elegantly provides a solution to the faint young Sun paradox through the UV shielding of reduced gases while itself creating a large source of biological precursors.

The work presented here demonstrates the ability of a Titan-like haze to act as an effective UV shield on the early Earth. The haze as described in this work would surely allow the buildup photolytically unstable reduced gases in the atmosphere, but the extent to which the constituents and chemistry of the atmosphere would change is not precisely known. A complete analysis of the Archean haze and its impacts on the faint young Sun paradox requires the use of a fully coupled climate model linking atmospheric chemistry, aerosol microphysics, and radiative transfer.

References and Notes

1. C. Sagan, G. Mullen, *Science* **177**, 52 (1972).
2. P. Knauth, D. R. Lowe, *Geol. Soc. Am. Bull.* **115**, 566 (2003).
3. N. H. Sleep, A. M. Hessler, *Earth Planet. Sci. Lett.* **241**, 594 (2006).
4. R. Rye, P. H. Kuo, H. D. Holland, *Nature* **378**, 603 (1995).
5. A. A. Pavlov, L. L. Brown, J. F. Kasting, *J. Geophys. Res.* **106**, 23267 (2001).
6. J. D. Haqq-Misra, S. D. Domagal-Goldman, P. J. Kasting, J. F. Kasting, *Astrobiology* **8**, 1127 (2008).
7. M. G. Tomasko et al., *Planet. Space Sci.* **56**, 669 (2008).
8. H. L. DeWitt et al., *Astrobiology* **9**, 447 (2009).
9. C. Sagan, C. Chyba, *Science* **276**, 1217 (1997).
10. A. Bar-Mun, I. Kleinfeld, E. Ganor, *J. Geophys. Res.* **93**, 8383 (1988).
11. Details of the fractal microphysical model, along with additional results, are available as supporting online material (SOM) on Science Online.
12. M. Cabane, P. Rannou, E. Chassefiere, G. Israel, *Planet. Space Sci.* **41**, 257 (1993).
13. P. Rannou et al., *Icarus* **118**, 355 (1995).
14. P. Rannou, M. Cabane, R. Botet, E. Chassefiere, *J. Geophys. Res.* **102**, 10997 (1997).
15. P. Rannou, C. P. McKay, R. D. Lorenz, *Planet. Space Sci.* **51**, 963 (2003).
16. B. B. Mandelbrot, *Fractals: Form, Chance and Dimension* (Freeman, San Francisco, 1977).
17. R. Botet, P. Rannou, M. Cabane, *Appl. Opt.* **36**, 8791 (1997).
18. The mean field approximation for Mie scattering by fractal aggregates of identical spheres has been validated against exact solutions for scattering by silica aggregates with $D_f = 2$, $\eta_{mon} = 64$, a monomer optical size parameter of 3.9, and a complex refractive index of $1.4 + 0.0001i$. Large absorbing aggregates such as those found in this study have not been rigorously validated; however, the mean field fractal optical model has been successful in replicating the optical properties of Titan fractal hazes (24, 25). Additional discussion of the validity of the fractal optical model is included in the SOM.
19. B. N. Khare et al., *Icarus* **60**, 127 (1984).
20. Sensitivity studies indicate that the UV-shielding characteristics of a fractal aggregate haze are largely independent of the chosen refractive indices. Details are available in the SOM.
21. M. Pham, J.-F. Muller, G. P. Brasseur, C. Granier, G. Megie, *J. Geophys. Res.* **100**, 26061 (1995).
22. The effective optical depth is defined as $\tau_{eff} = (1 - \omega/2 - g\omega/2)\tau_{ext}$, where ω is the single scattering albedo, g is the asymmetry parameter, and τ_{ext} is the extinction optical depth. The effective optical depth estimates the total attenuation of downwelling radiation when both absorption and backscatter are considered.
23. J. Farquhar, J. Savarino, S. Airieau, M. H. Thiemens, *J. Geophys. Res.* **106**, 32829 (2001).
24. S. L. Miller, *Science* **117**, 528 (1953).
25. W. R. Kuhn, S. K. Atreya, *Icarus* **37**, 207 (1979).
26. J. F. Kasting, *J. Geophys. Res.* **87**, 3091 (1982).
27. K. J. Zahnle, *J. Geophys. Res.* **91**, 2819 (1986).
28. J. F. Kasting, D. H. Egger, S. P. Raeburn, *J. Geol.* **101**, 245 (1993).
29. This work was supported by NASA grant NNX08AG93G.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5983/1266/DC1

Materials and Methods

SOM Text

Figs. S1 to S14

Table S1

References

13 October 2009; accepted 7 April 2010

10.1126/science.1183260

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1269>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1269/DC1>

This article **cites 20 articles**, 4 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/328/5983/1269#otherarticles>

This article has been **cited by 1** articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/328/5983/1269#otherarticles>

This article appears in the following **subject collections**:

Evolution

<http://www.sciencemag.org/cgi/collection/evolution>

Natural and Sexual Selection in a Wild Insect Population

R. Rodríguez-Muñoz,¹ A. Bretman,^{1,2} J. Slate,³ C. A. Walling,⁴ T. Tregenza^{1*}

The understanding of natural and sexual selection requires both field and laboratory studies to exploit the advantages and avoid the disadvantages of each approach. However, studies have tended to be polarized among the types of organisms studied, with vertebrates studied in the field and invertebrates in the lab. We used video monitoring combined with DNA profiling of all of the members of a wild population of field crickets across two generations to capture the factors predicting the reproductive success of males and females. The factors that predict a male's success in gaining mates differ from those that predict how many offspring he has. We confirm the fundamental prediction that males vary more in their reproductive success than females, and we find that females as well as males leave more offspring when they mate with more partners.

Insects are of fundamental importance to terrestrial ecosystems but are underrepresented in studies that aim to understand how natural and sexual selection drive evolution in wild populations. Although poorly understood in their natural habitats, crickets have become an important laboratory model system, revealing complex forms of sexual selection whereby females choose between males according to their songs (1), males fight (2), females manipulate sperm from several males to favor unrelated males (3, 4), and females lay eggs faster when mated to dominant males (2). However, although we now have many insights into the behavior and physiology of crickets in the laboratory, we have almost no idea how important these various aspects are in the insects' natural habitat. This discrepancy is a cause for concern: Laboratory situations remove some sources of selection that may be very important in wild populations and may create new pressures; for instance, it may be that males that sing more get more mates in the lab, but in the field such males may die younger.

Univoltine flightless field crickets, *Gryllus campestris*, hatch from eggs in early summer. Nymphs build burrows among the grass and spend the winter underground, emerging in spring to undergo one or two final molts to adulthood. Both sexes are highly territorial and spend the vast majority of their time in the immediate vicinity of a burrow entrance. A few days after becoming adults, males start to sing, and both males and females start moving frequently from one burrow to another in search of mates. To identify selective pressures affecting behavior and to observe how behavior is correlated with fitness, we built a network of 64 motion-sensitive, infrared-equipped

video cameras allowing us to monitor occupied burrows 24 hours a day throughout the breeding season. We tagged every newly emerged adult with a unique code to analyze their lives and behaviors, including mating partners, how long particular males and females spent together, the time that each male spent singing calling songs to attract females, and the fights that almost invariably occur when a male approaches a burrow occupied by another male. We used these fights to score males as either dominant or subordinate, reflecting the proportion of fights that he won (5). Although females never share burrows, they are only very rarely involved in aggressive interactions. Females visit or receive visits from neighboring males and frequently remain with a male for hours or days, sharing his burrow and mating repeatedly. From our videos, we inferred adult life span as the time from the observed emergence to the point when a cricket was either seen to be killed by a predator or was no longer found at any burrow.

We observed that females began mating a few days after becoming adults and laid eggs directly into the ground throughout the breeding season (burrows are narrow, so molting and mating take place just outside and are easily observed). The crickets in the field in the second year of our ob-

servations are therefore the offspring of the adults from the previous summer. Populations may experience some migration, but this is likely to be very limited in our study population. The meadow is relatively isolated, being surrounded by little suitable habitat, and the observed immigration rates of adults are low; therefore we had high success in assigning parentage within the population (5). All of these factors indicate that it is unlikely that substantial numbers of adult offspring were missed because of emigration. Lifetime reproductive success (LRS) was therefore inferred from the assignment of parentage from parents in 2006 to offspring in 2007 through the genotyping of all adults at 11 microsatellite loci.

A key prediction of the theory of sexual selection (6–8), assuming conventional sex roles and an even sex ratio, posits that males should have greater variance in LRS than females do. This prediction has been supported in a small number of studies of wild vertebrates [for example, (9)] and in laboratory experiments [although the lack of ecological context has led to debates over their relevance (10)]. Most studies of the cost and benefits of mates and matings in insects have been performed in the laboratory (11–13), and the only examination in the wild was of reproductive success estimated via the time female damselflies spent laying eggs after mating to a particular male (14). We directly examined both the number of mates that each individual had (controlling for differences in observational effort) and the number of descendants they left in the next generation. The sex ratio was very close to even, which constrained the means to be the same for both sexes. As expected, we found that mean numbers of mates per day [males = 0.27 (0.40 SD), females = 0.25 (0.32 SD)] and of offspring surviving to adulthood [males = 1.92 (3.66 SD), females = 1.79 (2.46 SD)] were, respectively, very similar. The small differences we observed were attributed to imperfections in observational data and in parentage assignment.

The opportunity for selection can be estimated by comparing variances or coefficients of variation (15). We examined with a randomization-

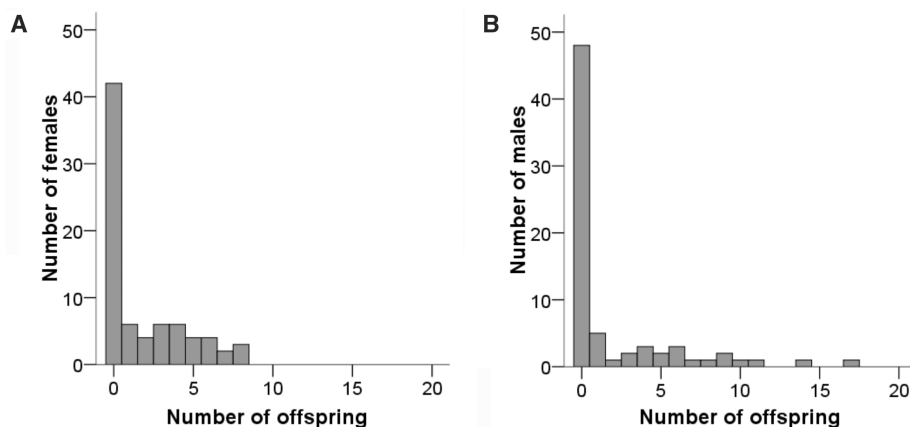


Fig. 1. Number of adult offspring per individual. Frequencies for (A) females and (B) males. Males have significantly greater variance in offspring number relative to females.

¹Centre for Ecology and Conservation, School of Biosciences, University of Exeter, Cornwall Campus, Penryn TR10 EZ, UK.

²School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK. ³Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK. ⁴Institute of Evolutionary Biology, School of Biological Sciences, Ashworth Laboratories, King's Buildings, University of Edinburgh, Edinburgh EH9 3JT, UK.

*To whom correspondence should be addressed. E-mail: t.tregenza@exeter.ac.uk

based test (5) the prediction that males should show more variation than females, and we found no difference between the sexes in variance in the number of mates ($P = 0.39$) but significantly more variance in the number of offspring produced by males ($P = 0.033$) (Fig. 1) (5). Similar results were found when controlling for the small differences in means between males and females (number of mates, $P = 0.39$; number of offspring, $P = 0.028$). It is, however, striking that although males varied more than females, the overall patterns were similar, with many females failing to leave any descendants.

We found that both sexes benefit from multiple mates. Females may use a single ejaculate over their reproductive lifetime, raising questions about why they mate with more than one male (16). Examination of traits expressed by both sexes indicated that individuals with a higher number of mates (standardized for monitoring effort) had a higher number of offspring for both male parents (Spearman's $\rho = 0.53$, $n = 47$ individuals, $P < 0.0001$) and female parents (Spearman's $\rho = 0.37$, $n = 55$, $P = 0.005$). This suggests that the factors affecting the number of offspring produced by males are the same as those affecting the number of offspring produced by females. Because there are numerous other correlations between traits, we also examined whether the number of mates was correlated with other behavioral and life-history traits by using a generalized linear model (5). This approach indicated that there were no interactions between sex and number of mates affecting LRS, nor any other significant interactions. Individuals of both sexes that were either larger, longer-lived, or had more mates had significantly higher LRS (analysis of deviance: size deviance = 4.58, $F_{1, 99} = 4.58$, $P = 0.035$; longevity deviance = 117.14, $F_{1, 99} = 38.74$, $P < 0.0001$; number of mates deviance = 44.47, $F_{1, 99} = 14.71$, $P = 0.0002$) (Fig. 2). This demonstrates that not only do males increase their reproductive success

through an increased number of mates, but females do better by mating with multiple partners, too. Although there is greater variance in male reproductive success than in female reproductive success, this increased variance is not due to different effects of mate number between the sexes.

Traits that confer success in gaining mates may differ from those that predict reproductive success because of differences among males in post-mating fertilization success and the viability of offspring (17). We counted the number of offspring an individual had in the following generation to measure directly the reproductive success of our field crickets. We found differences between predictors of mating success and predictors of reproductive success. By comparing generalized linear models predicting mating success (measured as the number of females with which each male was observed to mate) with lifetime reproductive success (number of offspring surviving to adulthood), we identified factors that predicted the reproductive success of an individual male. Mating success was predicted by dominance and an interaction between size and singing activity. Offspring number was also predicted by interactions between size and singing, but there was an additional interaction between longevity and singing and no significant effect of dominance. Contrary to expectations from lab studies that show that dominant males can monopolize mating access to females (18), who also prefer the odor of dominant males (19), we found that dominant males had fewer mates than did subordinate males (analysis of deviance: dominance deviance = 2.22, $F_{1, 24} = 9.58$, $P = 0.004$) (Fig. 3A). This result is unexpected but reflects an ambiguous role for dominance in predicting mating success across species (20). For smaller males, the amount of singing was strongly correlated with the number of mates they obtained, whereas for larger males, singing activity was not associated with gaining more mates (size $\times$ singing deviance = 0.89, $F_{1, 23} =$

4.53, $P = 0.044$) (Fig. 3B). Similarly, for small males, singing effort affected the number of offspring (analysis of deviance: size $\times$ singing deviance = 25.93, $F_{1, 41} = 8.12$, $P = 0.007$) (Fig. 3C). In addition, short-lived males had more offspring when they sang more, whereas reproductive success in long-lived males was not dependent on a high rate of singing (longevity $\times$ singing deviance = 13.91, $F_{1, 41} = 4.35$, $P = 0.043$) (Fig. 3D).

These interactions between naturally and sexually selected traits affecting different measures of reproductive success indicate that the benefits of sexually selected traits may vary according to other aspects of an individual's phenotype. The costs of sexually selected traits are expected to be lower in individuals of higher phenotypic condition, which in turn reflects the overall genetic quality (21). If males with high genetic quality are able to achieve a large size at adult emergence, they would be predicted to sing more. However, we observed no correlation between size and singing activity within males (Spearman's $\rho = 0.185$, $P = 0.2$, $n = 47$). Our results suggest that in this population, either size is not a reliable indicator of condition or male singing activity is not condition-dependent, despite its metabolic costs (22) and association with increased rates of parasitism (23) and predation (24). Size and singing activity are individually correlated with reproductive success, but the fact that smaller males benefit more from singing is the opposite of what we would expect if larger males are deemed to be those with the higher condition. This suggests that adult size is not an appropriate proxy for phenotypic condition in these animals and possibly in other insects.

Longevity is a fundamentally different trait from size, because it is not fixed at emergence to adulthood and hence can continue to be affected by other traits. The interaction between singing and longevity affecting reproductive success occurs because daily singing effort has a major effect on offspring number in shorter-lived males

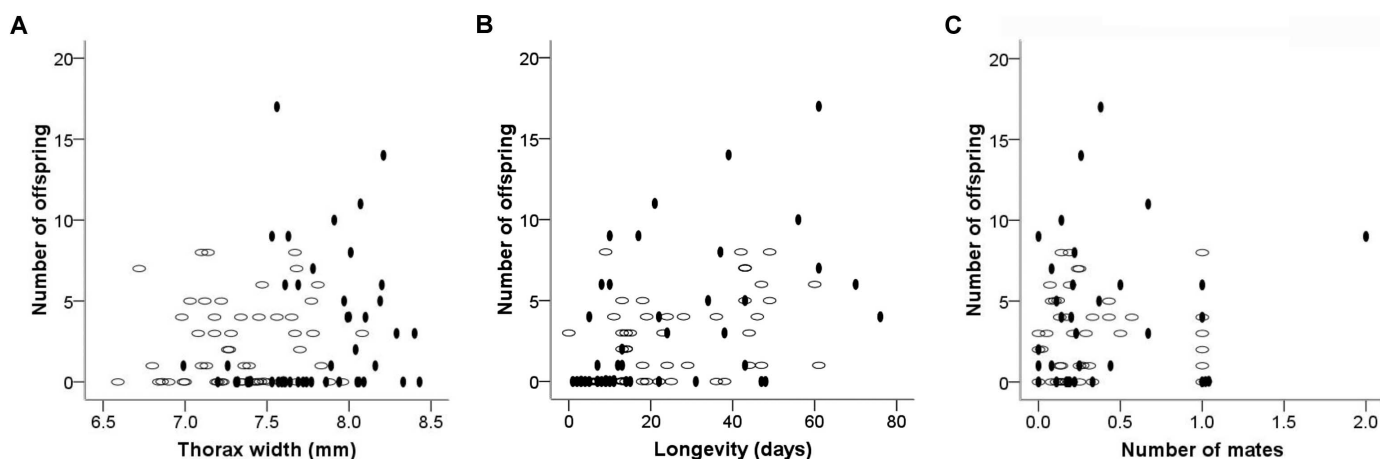


Fig. 2. Determinants of male and female reproductive success measured as number of adult offspring in the following generation. Females are indicated by open symbols. (A) Body size, (B) longevity (days), and (C) number of mates per day. To aid in distinguishing data, where data points overlap, a small increment (0.02 to x and y coordinates) has been added.

Also, in (C), a number of points overlap at 0, 0 (16 males and 12 females) and would obscure further data, so have not received an increment. Larger body size, greater longevity, and a higher number of mates are independently associated with increased reproductive success in both males and females.

but has no discernible effect in longer-lived males (Fig. 3D). This may be because if males live a long time, singing effort can be reduced once they have attracted a female to their burrow, whereas the population of short-lived males includes more individuals that are in poor condition and die young without mating or singing very much. In the laboratory, male crickets kept on a high-protein diet sang more, but lived for a shorter time (25). In the wild population, there was a strong positive correlation between daily singing effort and life span (Spearman's $\rho = 0.54$, $P < 0.0001$, $n = 47$), most likely indicating that those that sang more were of higher quality.

Our findings confirm the basic prediction that male reproductive success, while being constrained to be equal to that of females, is likely to vary more. Bateman's prediction (7) that this variation is due to the potentially higher mating rate of males does not appear to be borne out; variance in the number of mates a male had was no greater

than that of females. Rather, it appears that in these insects, some males gain a disproportionately large share of the offspring in the following generation through either greater success in postcopulatory sexual selection or greater viability and survival of their offspring.

Both sexes have higher LRSs when they have more mates, and there is no interaction between sex and number of mates that affects LRS. This demonstrates that polyandrous females have more offspring in the next generation, supporting previous laboratory experiments (26). Because polyandrous females tend to have more matings as well as more mates, it is still unknown whether these benefits accrue from some direct source such as ejaculate components or a need to replenish sperm stores, or whether they are the result of genetic benefits to the offspring of polyandrous females. It may be that polyandrous females can increase offspring fitness by preferentially fertilizing their eggs with sperm from unrelated males

(3) and that, by mating with multiple males, females increase their chances of producing offspring with unrelated males.

The number of mates was a strong predictor of the number of descendants that both males and females left in the next generation. Similarly, traits associated with having several mates, such as male size and singing activity, were also associated with LRS, indicating that, in general, mating success is likely to perform quite well as a surrogate for overall reproductive success. However, it is clear that if we wish to understand selection on individual traits in natural populations, careful consideration must be given to how we measure reproductive success: Some traits that affected an individual's number of mates failed to predict LRS, and other traits associated with high LRS were not good predictors of mating success. Furthermore, interactions between naturally and sexually selected traits affecting both mating and reproductive success indicate that studying a single trait in isolation may be misleading.

Our system bridges the divide between laboratory and field studies in evolutionary biology and indicates that, with the above caveats, conclusions drawn from laboratory studies on crickets and, most likely, other insects as well are generally consistent with studies in the wild. We demonstrate here that the combination of video technology and genetic parentage assignment means that tracing reproductive success in wild invertebrates is no longer impractical and that we can now conduct quantitative and functional genetic studies in natural populations.

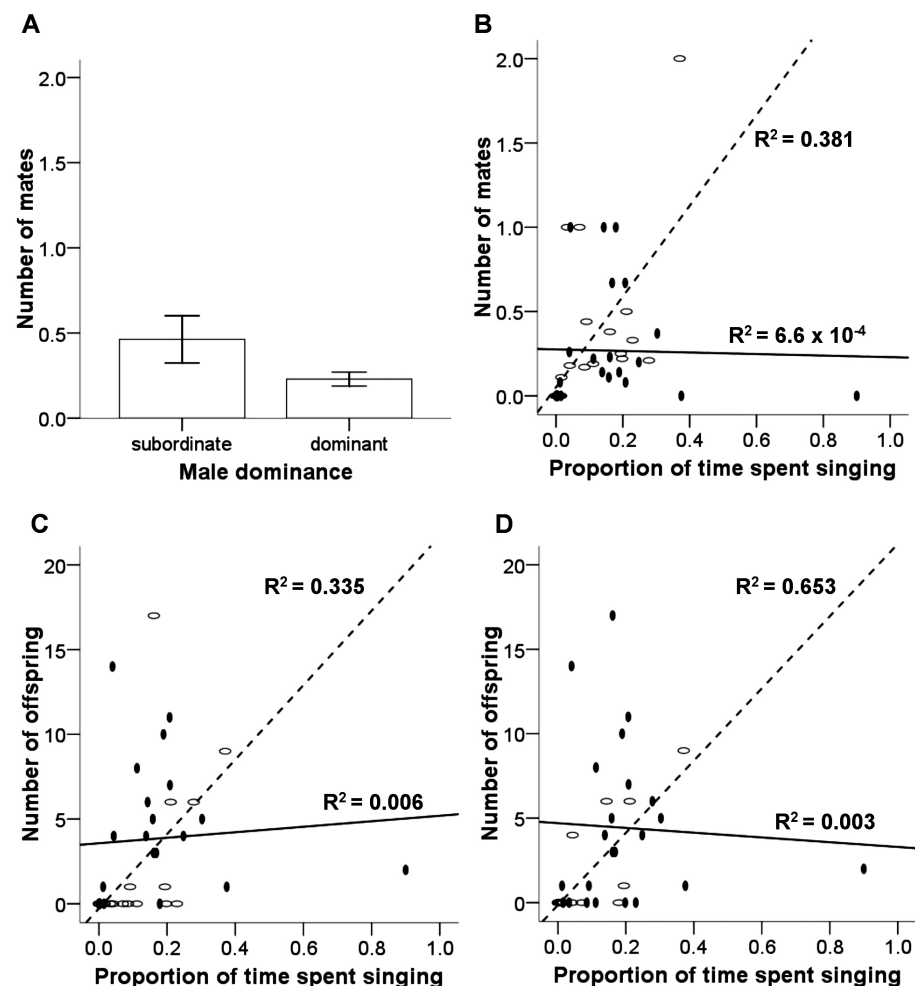


Fig. 3. Male sexual traits and mating and reproductive success. The effect on number of mates of (A) dominance and (B) the interaction between size and amount of singing, and the effect on number of adult offspring of (C) the interaction between size and the amount of singing and (D) the interaction between longevity and the amount of singing. For illustrative purposes only, the 47 males were split into two groups by size (24 smaller and 23 larger males) or by longevity (24 shorter-lived and 23 longer-lived males). Open symbols and hatched lines indicate smaller (B) and (C) or shorter-lived (D) males. A number of data points overlap at 0, 0: nine smaller males and five larger in (B) and (C), and 14 short-lived males in (D).

References and Notes

1. M. Zuk, L. Simmons, in *The Evolution of Mating Systems in Insects and Arachnids*, J. C. Choe, B. J. Crespi, Eds. (Cambridge Univ. Press, Cambridge, 1997), pp. 89–109.
2. A. Bretman, R. Rodríguez-Muñoz, T. Tregenza, *Biol. Lett.* **2**, 409 (2006).
3. A. Bretman, D. Newcombe, T. Tregenza, *Mol. Ecol.* **18**, 3340 (2009).
4. A. Bretman, N. Wedell, T. Tregenza, *Proc. Biol. Sci.* **271**, 159 (2004).
5. Supporting material is available on Science Online.
6. C. R. Darwin, *The Descent of Man and Selection in Relation to Sex* (John Murray, London, 1871).
7. A. J. Bateman, *Heredity* **2**, 349 (1948).
8. S. M. Shuster, M. J. Wade, *Mating Systems and Strategies* (Princeton Univ. Press, Princeton, NJ, 2003).
9. B. F. Snyder, P. A. Gowaty, *Evolution* **61**, 2457 (2007).
10. T. H. Clutton-Brock, *Reproductive Success* (Univ. of Chicago Press, Chicago, 1988).
11. T. Bilde, A. Foged, N. Schilling, G. Arnqvist, *Science* **324**, 1705 (2009).
12. T. Chapman, L. F. Liddle, J. M. Kalb, M. F. Wolfner, L. Partridge, *Nature* **373**, 241 (1995).
13. G. A. Schwarzenbach, D. J. Hosken, P. I. Ward, *J. Evol. Biol.* **18**, 455 (2005).
14. O. M. Fincke, *Evolution* **40**, 791 (1986).
15. H. Kokko, A. Mackenzie, J. D. Reynolds, J. Lindström, W. J. Sutherland, *Am. Nat.* **154**, 358 (1999).
16. M. D. Jennions, M. Petrie, *Biol. Rev. Camb. Philos. Soc.* **75**, 21 (2000).
17. W. G. Eberhard, *Female Control: Sexual Selection by Cryptic Female Choice* (Princeton Univ. Press, Princeton, NJ, 1996).
18. N. Wedell, T. Tregenza, *Evolution* **53**, 620 (1999).
19. R. Kortet, A. Hedrick, *Behav. Ecol. Sociobiol.* **59**, 77 (2005).

20. A. Qvarnström, E. Forsgren, *Trends Ecol. Evol.* **13**, 498 (1998).
21. L. Rowe, D. Houle, *Proc. Biol. Sci.* **263**, 1415 (1996).
22. M. A. Hack, J. *Insect Behav.* **11**, 853 (1998).
23. M. Zuk, L. W. Simmons, L. Cupp, *Behav. Ecol. Sociobiol.* **33**, 339 (1993).
24. W. J. Bailey, S. Haythornthwaite, *J. Zool. (London)* **244**, 505 (1998).
25. J. Hunt *et al.*, *Nature* **432**, 1024 (2004).
26. T. Tregenza, N. Wedell, *Evolution* **52**, 1726 (1998).
27. This work was supported by the Natural Environment Research Council (NERC) (grant NE/E005403/1), the NERC Sheffield Biomolecular Analysis Facility, the Leverhulme Trust, and a Royal Society Fellowship to T.T. We thank L. Rodríguez and M. C. Muñoz for providing access and facilities at the study site; J. M. López, C. López, E. González, C. R. del Valle, and F. González for assistance with the field work; and D. J. Hodgson and A. Wetherelt for assistance with data analysis and surveying.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5983/1269/DC1
Materials and Methods
Figs. S1 and S2
Table S1
References
Movies S1 to S8

9 February 2010; accepted 12 April 2010
10.1126/science.1188102

Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance

Jesse D. Bloom, Lizhi Ian Gong, David Baltimore*

The His²⁷⁴→Tyr²⁷⁴ (H274Y) mutation confers oseltamivir resistance on N1 influenza neuraminidase but had long been thought to compromise viral fitness. However, beginning in 2007–2008, viruses containing H274Y rapidly became predominant among human seasonal H1N1 isolates. We show that H274Y decreases the amount of neuraminidase that reaches the cell surface and that this defect can be counteracted by secondary mutations that also restore viral fitness. Two such mutations occurred in seasonal H1N1 shortly before the widespread appearance of H274Y. The evolution of oseltamivir resistance was therefore enabled by “permissive” mutations that allowed the virus to tolerate subsequent occurrences of H274Y. An understanding of this process may provide a basis for predicting the evolution of oseltamivir resistance in other influenza strains.

Influenza A is a respiratory virus that causes annual epidemics and occasional pandemics, of which the worst on record killed in excess

of 20 million people worldwide (1). One of the main defenses against influenza is the antiviral drug oseltamivir (Tamiflu, F. Hoffmann-La Roche, Incorporated) (2), and over 200 million doses have been stockpiled worldwide (3). Oseltamivir binds in the active site of the neuraminidase (NA) enzyme expressed on the virion surface, preventing it from cleaving sialic acid moieties that can be bound by the viral hemagglutinin

protein (2). This lack of NA activity inhibits the release of newly formed virions from infected cells (4), as well as causing viral aggregation (4), reducing infectivity (5, 6), and limiting the ability of viruses to penetrate mucus found in the airways (6).

During clinical testing of oseltamivir, a small fraction of human participants who were infected with the seasonal human H1N1 influenza strain A/Texas/36/1991 (TX91) and then treated with oseltamivir eventually shed resistant viruses (7). These viruses carried a mutation of histidine to tyrosine at NA residue 274 (H274Y), which is found near but not directly in the substrate-binding pocket (8). This mutation causes subtle structural alterations that weaken oseltamivir binding (8, 9). However, TX91 viruses with H274Y were attenuated in tissue culture, mice, and ferrets (10). H274Y also impaired the growth of the H1N1 lab strain A/WSN/33 (WSN) in tissue culture (11) and the infectivity of the seasonal H1N1 strain A/New Caledonia/20/1999 (NC99) in ferrets (12). These studies led to the conclusion that “[v]irus carrying a H274Y mutation is unlikely to be of clinical consequence (10).”

This conclusion held sway from the introduction of oseltamivir as a drug in 1999 until the 2007–2008 influenza season, when oseltamivir-

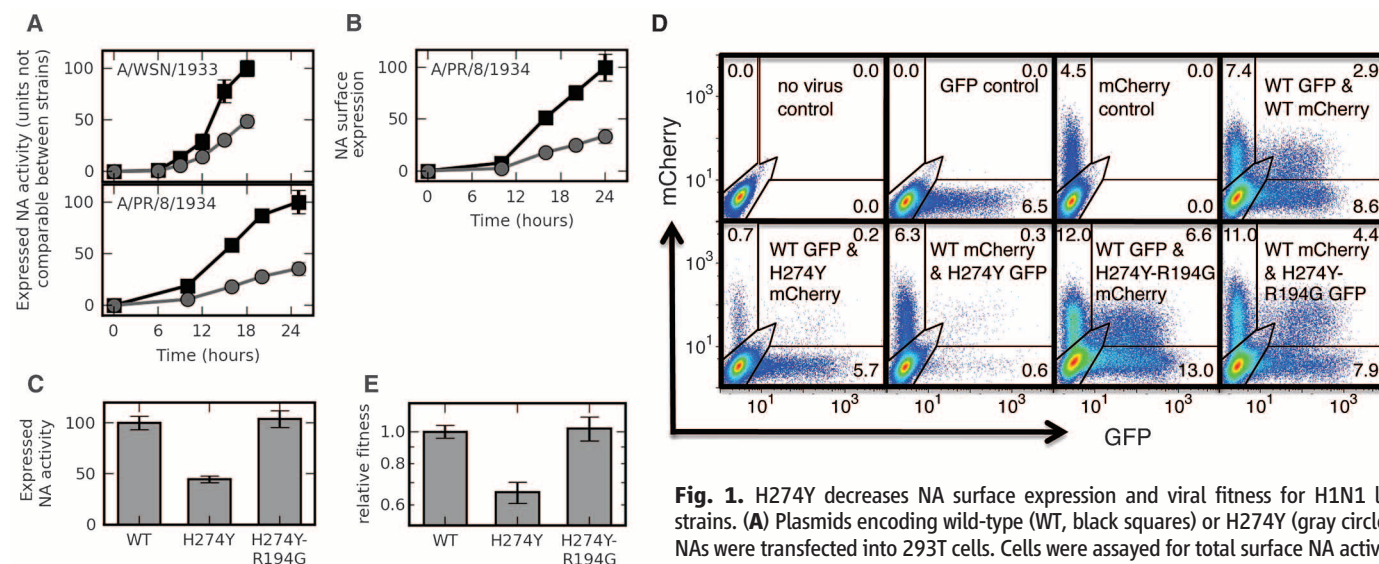


Fig. 1. H274Y decreases NA surface expression and viral fitness for H1N1 lab strains. **(A)** Plasmids encoding wild-type (WT, black squares) or H274Y (gray circles) NAs were transfected into 293T cells. Cells were assayed for total surface NA activity in a nonlysing buffer. **(B)** The same cells were assayed by flow cytometry for mean mCherry expression. **(C)** Surface activity of WT, H274Y, or H274Y-R194G WSN NA at 18 hours. All four graphs show mean $\pm$ SEM for at least three independent transfections. **(D)** An estimated 50 infectious particles of each color of WSN-PB1flank-eGFP and/or mCherry carrying the indicated NA mutations were infected into 2.5×10^5 A549-CMV-PB1 cells. After 46 to 50 hours, the cells were analyzed by flow cytometry to quantify virus growth. Numbers give percentages of cells in each gate; one representative plot is shown for each virus pair. **(E)** Relative fitness (ratio of Malthusian growth parameters) with respect to WT calculated from at least eight replicates. Shown are geometric means and 95% confidence intervals.

surface NA expression using a monoclonal antibody against the PR8 tetramer. **(C)** Surface activity of WT, H274Y, or H274Y-R194G WSN NA at 18 hours. All four graphs show mean $\pm$ SEM for at least three independent transfections. **(D)** An estimated 50 infectious particles of each color of WSN-PB1flank-eGFP and/or mCherry carrying the indicated NA mutations were infected into 2.5×10^5 A549-CMV-PB1 cells. After 46 to 50 hours, the cells were analyzed by flow cytometry to quantify virus growth. Numbers give percentages of cells in each gate; one representative plot is shown for each virus pair. **(E)** Relative fitness (ratio of Malthusian growth parameters) with respect to WT calculated from at least eight replicates. Shown are geometric means and 95% confidence intervals.

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
<http://www.sciencemag.org/cgi/content/full/328/5983/1272>

Supporting Online Material can be found at:
<http://www.sciencemag.org/cgi/content/full/328/5983/1272/DC1>

This article **cites 32 articles**, 14 of which can be accessed for free:
<http://www.sciencemag.org/cgi/content/full/328/5983/1272#otherarticles>

This article has been **cited by 1 articles** hosted by HighWire Press; see:
<http://www.sciencemag.org/cgi/content/full/328/5983/1272#otherarticles>

This article appears in the following **subject collections**:
Virology
<http://www.sciencemag.org/cgi/collection/virology>

resistant H274Y seasonal H1N1 viruses appeared worldwide (13). Within a year, H274Y was present in most seasonal H1N1 (13). These

viruses showed no obvious attenuation relative to earlier viruses carrying tyrosine at position 274 (14).

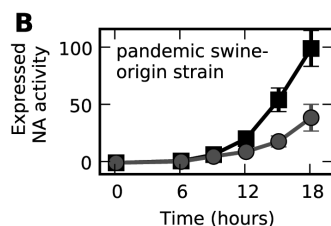
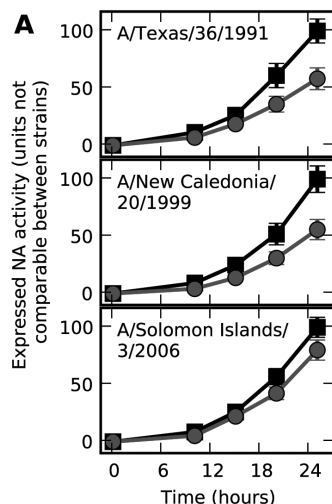


Fig. 2. Total surface-expressed activity for WT (black squares) and H274Y (gray circles) NAs from (A) three seasonal human H1N1 strains and (B) the pandemic swine-origin A/California/4/2009 (H1N1) strain. All graphs show mean $\pm$ SEM for at least three replicates.

We hypothesized that the evolution of oseltamivir resistance was enabled by permissive mutations that alleviated the deleterious effects of subsequent occurrences of H274Y. Earlier studies have failed to find large effects of H274Y on substrate affinity or catalytic rate in otherwise isogenic NAs (9–11). We therefore further hypothesized that these permissive mutations buffered previously unobserved deficiencies in NA folding or stability caused by H274Y. This hypothesis was motivated by the observation that functionally adaptive mutations often come at a cost to protein stability or folding and that mutations that bolster these properties can therefore promote evolvability (15).

Preliminary experiments uncovered no obvious differences in the stabilities of wild-type and H274Y NAs to irreversible thermal denaturation (fig. S1), so we examined whether H274Y hampered the efficiency with which folded NA reached the cell surface. Plasmids encoding wild-type or H274Y NA were transfected into 293T

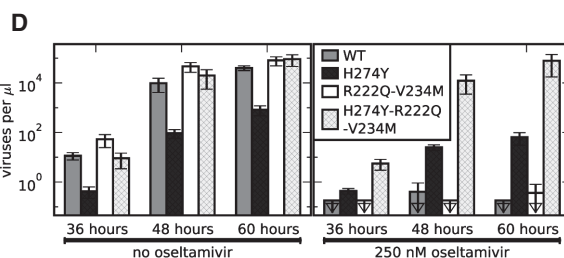
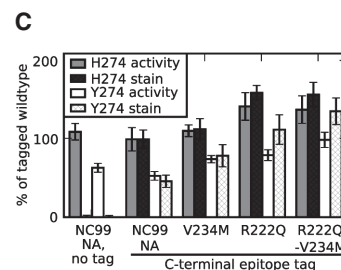
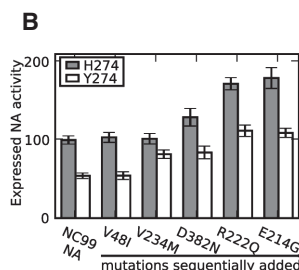
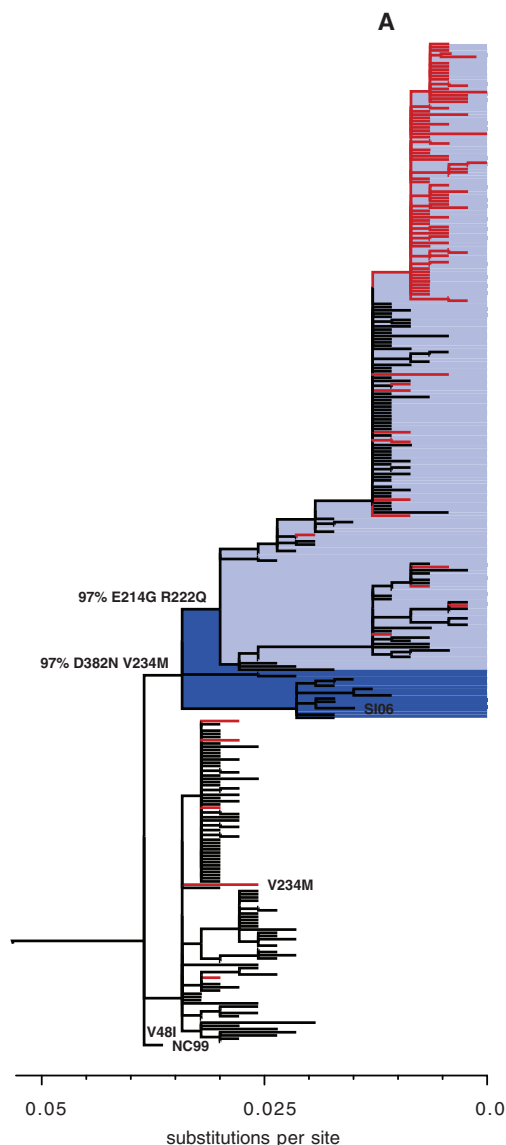


Fig. 3. Identification of mutations that buffer defects in NA surface expression and viral fitness caused by H274Y in seasonal human H1N1. (A) Maximum-likelihood phylogenetic tree showing NC99 and all seasonal H1N1 strains from 2006 or later. Based on a maximum-likelihood reconstruction, branches with H274Y are colored red, whereas those without are colored black. Indicated are the first five mutations on the path from NC99 to the large group of H274Y viruses. Bootstrap percentages are shown for key nodes. Branches with V234M are shaded dark blue, and those with both V234M and R222Q are shaded light blue. One of the isolated H274Y sequences contained an independent appearance of V234M as indicated. The NC99 and SI06 sequences are labeled. (B) Surface-expressed activity at 20 hours for NC99 NAs with sequential addition of the five reconstructed mutations, both with and without H274Y. (C) Surface-expressed activity and mean staining intensity at 20 hours for NC99 NAs containing a C-terminal epitope tag. (D) An estimated 50 infectious particles of PB1flank-eGFP with the indicated NC99 NAs and the remaining genes from TX91 were infected into 2.5×10^5 MDCK-SIAT1-CMV-PB1 cells, and the viral titer in the supernatant was determined at the indicated times. All graphs show mean $\pm$ SEM for at least three replicates.

cells, and the total NA activity expressed on the cell surface was measured by using a fluorogenic substrate (16). Cells transfected with wild-type NA expressed at least twofold more total surface activity than those transfected with H274Y, both for NAs from WSN and from another H1N1 lab strain, A/PR/8/34 (PR8) (Fig. 1A). Previous studies found little effect of H274Y on the catalytic activity per unit enzyme (9–11), suggesting that the differences are due to fewer NAs at the cell surface. To test this, we used a monoclonal antibody to quantify the amount of PR8 NA on the surface of transfected cells (Fig. 1B). Surface expression closely corresponded to total surface activity, confirming that H274Y decreases the amount of NA that is inserted into the membrane.

An evolutionarily important feature of mutations that interfere with protein folding is that they can frequently be compensated for by other mutations at sites distant in the three-dimensional structure (17, 18). We used a computational method (19) to predict that Arg¹⁹⁴→Gly¹⁹⁴ (R194G), which occurs over 20 Å from H274Y in the folded structure (fig. S2), might have this compensatory effect on WSN NA. Introducing R194G into H274Y WSN NA restored the total surface-expressed activity to about wild-type levels (Fig. 1C).

To examine whether a mutation that restored NA surface expression would also restore viral fitness, we created WSN viruses possessing PB1 segments in which most of the coding sequence was replaced by a gene encoding either an enhanced green or a red fluorescent protein (eGFP or mCherry) (fig. S3). Because these segments retained the noncoding and 80 terminal coding nucleotides, they were packaged efficiently and stably into virions (20, 21) (fig. S3). These PB1flank-eGFP/mCherry viruses grew to high titers in cells constitutively expressing the PB1 protein. We created variants of both colors carrying wild-type, H274Y, or H274Y-R194G NAs and performed viral growth assays by adding about 50 infectious particles of each color to 2.5×10^5 A549 human lung carcinoma cells that expressed PB1 protein. After 46 to 50 hours, flow cytometry was used to quantify the cells infected by each color virus (Fig. 1D). WSN viruses with H274Y were strongly attenuated relative to wild type (Fig. 1E), in agreement with earlier work (11). However, H274Y virus also carrying R194G exhibited fitness indistinguishable from that of the wild type (Fig. 1E).

The foregoing results indicate that the defect caused by H274Y can be corrected by secondary mutations. To determine whether this actually occurred in human influenza, we examined NAs from seasonal H1N1. Introducing H274Y into TX91 and NC99 NAs caused an ~twofold decrease in total surface expressed activity (Fig. 2A); earlier studies have shown that H274Y attenuates these strains in ferrets (10, 12). But H274Y caused a much smaller decrease in the more recent A/Solomon Islands/3/2006 (SI06)

strain (Fig. 2A). We built a phylogenetic tree incorporating NC99 and all seasonal H1N1 NAs from 2006 or later (Fig. 3A) and traced along this tree from NC99 toward the large group of H274Y viruses, sequentially adding the first five reconstructed mutations in the order which maximum likelihood suggested that they occurred. Two of the reconstructed mutations had important effects on total surface-expressed NA activity (Fig. 3B). The first, Val²³⁴→Met²³⁴ (V234M) (which is found in SI06), decreased the magnitude of the defect caused by H274Y. The second, Arg²²²→Gln²²² (R222Q), increased total surface-expressed activity for NAs both with and without H274Y. Both of these mutations were retained in all members of the large group of H274Y sequences (Fig. 3A). Neither mutation is in close contact with H274 in the folded NA structure (fig. S2).

To confirm that V234M and R222Q increase the amount of NA that reaches the cell surface, we added a C-terminal epitope tag. This tag was readily stained by fluorescent antibodies and did not substantially interfere with NA folding or function, because it caused at most a slight (<10%) decrease in total surface-expressed activity (Fig. 3C). Both V234M and R222Q increased the amount of NA that reached the cell surface; the activity per enzyme was unaffected by V234M and slightly decreased by R222Q (Fig. 3C). The NC99 NA carrying R222Q and V234M in addition to H274Y expressed a total surface activity roughly equivalent to that of wild type.

We next constructed PB1flank-eGFP viruses carrying NC99 NAs with the relevant mutations. The non-NA segments for these viruses were derived from the closely related (>98% protein identity) TX91 strain. We engineered a cell line (MDCK-SIAT1) that expresses high levels of the sialic acid linkage preferred by human influenza (22) to also constitutively express PB1 protein. We infected 2.5×10^5 cells with an estimated 50 infectious particles and monitored viral growth. The H274Y virus was strongly attenuated, growing to ~100-fold lower titers than wild type (Fig. 3D). However, virus carrying H274Y in conjunction with R222Q and V234M grew to levels comparable to that of wild type, as did virus carrying just R222Q and V234M (Fig. 3D). In the presence of 250 nM oseltamivir, only viruses carrying H274Y grew to substantial titers, with the H274Y-R222Q-V234M triple mutant growing to over 100-fold higher titers than virus with H274Y alone (Fig. 3D). Therefore, by accumulating the permissive V234M and R222Q mutations and then acquiring H274Y, NC99 virus can become resistant to oseltamivir while maintaining high levels of fitness in the absence of drug.

It is interesting to speculate about what drove the spread of these mutations in seasonal H1N1. The permissive mutations could be the result of stochastic drift, genetic hitchhiking (23), or selection for antigenic change or tuning of the

NA/hemagglutinin balance (24). Once the permissive mutations were in place, H274Y could have been the product of any of these same forces or of direct selection for oseltamivir resistance. The fact that initial geographic prevalence of H274Y was not correlated with oseltamivir usage is generally viewed as evidence against direct selection for resistance (25), although the lack of persistent spatial structure in human influenza (26, 27) makes it difficult to ascribe selection pressures to specific geographic locations.

Regardless of the forces that eventually drove its spread, our results show that H274Y attenuates seasonal H1N1 unless there are permissive secondary mutations that maintain adequate surface NA expression. These mutations likely buffer defects in NA folding or transport that are caused by H274Y. The reduced fitness of H274Y viruses in the absence of such permissive mutations could be caused by insufficient NA to balance the receptor binding of hemagglutinin (24) or to stimulate proper virion assembly (28).

The evolution of H274Y in seasonal H1N1 adds to a growing collection of examples of the importance of permissive mutations in molecular evolution. These examples include the evolution of bacterial antibiotic resistance (29, 30), the escape of HIV from cytotoxic T lymphocytes (31), and the acquisition of new ligand specificity in vertebrate steroid receptors (32). An appealing aspect of our findings is that, in addition to elucidating the historic role of permissive mutations, they may also provide a basis for making informed predictions. The pandemic swine-origin 2009 A(H1N1) viruses that recently swept the globe remain mostly oseltamivir-sensitive, but scattered H274Y isolates have emerged (33). It remains unclear whether these isolates are a less-fit evolutionary dead ends or harbingers of a resistance mutation that will soon spread worldwide. Introducing H274Y into the swine-origin A/California/4/2009 NA causes a large drop in total surface-expressed activity (Fig. 2B). We therefore suggest that, unless these swine-origin viruses already express substantial excess NA, the long-term evolutionary potential of H274Y isolates might depend on secondary mutations that rescue NA surface expression.

References and Notes

1. N. P. Johnson, J. Mueller, *Bull. Hist. Med.* **76**, 105 (2002).
2. A. Moscona, *N. Engl. J. Med.* **353**, 1363 (2005).
3. A. L. Wan Po, P. Farndon, N. Palmer, *Emerg. Infect. Dis.* **15**, 1686 (2009).
4. C. Liu, M. C. Eichelberger, R. W. Compans, G. M. Air, *J. Virol.* **69**, 1099 (1995).
5. T. Suzuki et al., *J. Virol.* **79**, 11705 (2005).
6. M. N. Matrosovich, T. Y. Matrosovich, T. Gray, N. A. Roberts, H. D. Klenk, *J. Virol.* **78**, 12665 (2004).
7. L. V. Gubareva, L. Kaiser, M. N. Matrosovich, Y. Soo-Hoo, F. G. Hayden, *J. Infect. Dis.* **183**, 523 (2001).
8. R. J. Russell et al., *Nature* **443**, 45 (2006).
9. P. J. Collins et al., *Nature* **453**, 1258 (2008).
10. J. A. Ives et al., *Antiviral Res.* **55**, 307 (2002).
11. Y. Abed, N. Goyette, G. Boivin, *Antivir. Ther.* **9**, 577 (2004).

12. M. L. Herlocher *et al.*, *J. Infect. Dis.* **190**, 1627 (2004).
13. A. Moscona, *N. Engl. J. Med.* **360**, 953 (2009).
14. M. A. Rameix-Welti *et al.*, *PLoS Pathog.* **4**, e1000103 (2008).
15. J. D. Bloom, S. T. Labthavikul, C. R. Otey, F. H. Arnold, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5869 (2006).
16. Materials and methods are available as supporting material on Science Online.
17. A. R. Poteete, D. Rennell, S. E. Bouvier, L. W. Hardy, *Protein Sci.* **6**, 2418 (1997).
18. A. Mitraki, B. Fane, C. Haase-Pettingell, J. Sturtevant, J. King, *Science* **253**, 54 (1991).
19. J. D. Bloom, M. J. Glassman, *PLOS Comput. Biol.* **5**, e1000349 (2009).
20. Y. Muramoto *et al.*, *J. Virol.* **80**, 2318 (2006).
21. Y. Liang, Y. Hong, T. G. Parslow, *J. Virol.* **79**, 10348 (2005).
22. M. Matrosovich, T. Matrosovich, J. Carr, N. A. Roberts, H. D. Klenk, *J. Virol.* **77**, 8418 (2003).
23. L. Simonsen *et al.*, *Mol. Biol. Evol.* **24**, 1811 (2007).
24. R. Wagner, M. Matrosovich, H. D. Klenk, *Rev. Med. Virol.* **12**, 159 (2002).
25. P. Kramarz, D. Monnet, A. Nicoll, C. Yilmaz, B. Ciancio, *Euro Surveill.* **14**, 19112 (2009).
26. M. I. Nelson *et al.*, *PLoS Pathog.* **4**, e1000133 (2008).
27. A. Rambaut *et al.*, *Nature* **453**, 615 (2008).
28. D. P. Nayak, R. A. Balogun, H. Yamada, Z. H. Zhou, S. Barman, *Virus Res.* **143**, 147 (2009).
29. X. Wang, G. Minasov, B. K. Shoichet, *J. Mol. Biol.* **320**, 85 (2002).
30. S. J. Schrag, V. Perrot, B. R. Levin, *Proc. Biol. Sci.* **264**, 1287 (1997).
31. A. D. Kelleher *et al.*, *J. Exp. Med.* **193**, 375 (2001).
32. E. A. Ortlund, J. T. Bridgham, M. R. Redinbo, J. W. Thornton, *Science* **317**, 1544 (2007); published online 16 August 2007 (10.1126/science.1142819).
33. *Wkly. Epidemiol. Rec.* **85**, 37 (2010).
34. J.D.B. was supported by a Beckman Institute Postdoctoral Fellowship and the Irvington Institute Fellowship Program of the Cancer Research Institute. L.I.G. was supported by a Summer Undergraduate Research Fellowship from the California Institute of Technology. Oseltamivir carboxylate was kindly provided by Roche.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5983/1272/DC1

Materials and Methods

Figs. S1 to S3

Tables S1 and S2

2 February 2010; accepted 9 April 2010

10.1126/science.1187816

The Incidence of Fire in Amazonian Forests with Implications for REDD

Luiz E. O. C. Aragão^{1*†} and Yosio E. Shimabukuro^{2*}

Reducing emissions from deforestation and degradation (REDD) may curb carbon emissions, but the consequences for fire hazard are poorly understood. By analyzing satellite-derived deforestation and fire data from the Brazilian Amazon, we show that fire occurrence has increased in 59% of the area that has experienced reduced deforestation rates. Differences in fire frequencies across two land-use gradients reveal that fire-free land-management can substantially reduce fire incidence by as much as 69%. If sustainable fire-free land-management of deforested areas is not adopted in the REDD mechanism, then the carbon savings achieved by avoiding deforestation may be partially negated by increased emissions from fires.

Reducing emissions from deforestation and degradation (REDD) is one of the most cost-effective mitigation mechanisms (1) and could contribute to an emission reduction of 13 to 50 billion tons of carbon (Gt C) by 2100 (2). REDD is therefore a high-priority mechanism for mitigation of climate change within the United Nations Framework Convention on Climate Change (UNFCCC). The future of REDD implementation relies on forthcoming agreements to tackle the unresolved outcomes from the 15th Convention of the Parties, which took place in December 2009. These negotiations can largely influence the maintenance or replacement of the Kyoto Protocol beyond 2012 and the future of tropical forests. Policy-makers are considering a range of options for developing countries to receive financial incentives to reduce their deforestation rates (2). However, the efficacy of REDD as a climate change mitigation strategy depends, in particular, upon the stabilization of

deforestation and degradation of the world's largest rainforest, the Amazon.

Deforestation in the Brazilian Amazon (defined as clear cutting and conversion of the original forest cover to other land uses) has resulted in annual forest area loss of $18,918 \pm 1,576 \text{ km}^2$ (SEM) from 1998 to 2007, according to the National Institute for Space Research (INPE) in Brazil (3). It is estimated that this results in release of 0.28 (0.17 to 0.49) Gt C to the atmosphere annually (4), corresponding to 24% of the world's C emissions from land cover change [$1.15 (0.58\text{--}1.79) \text{ Gt C year}^{-1}$] (5). In principle, discontinuing ongoing deforestation through mechanisms such as REDD would protect a large fraction of the 86 Gt of the carbon stored in Amazonian forest biomass (6), which is equivalent to about a decade of global fossil fuel emissions to the atmosphere. However, there is a pressing need to consider the threat to forests posed by fire.

Fires following drought years are likely to release a similar amount of carbon as emissions from deliberate deforestation (7, 8). The combined effect of deliberate deforestation and forest fires has a similar magnitude to the natural annual carbon sink of 0.45 (0.3 to 0.6) Gt C estimated for undisturbed Amazonian forests (9). The higher probability of a drier Amazon in the 21st century predicted by some global circulation

models (10, 11), and consequent increasing drought intensity and frequency, may push Amazonia toward an amplified fire-prone system (12). Previous studies (13, 14) have shown an increase in fire occurrence following two large-scale Amazonian droughts (1998 and 2005). Changes in fire frequency could jeopardize the benefits achieved through REDD; however, despite its vital importance in this region, fire is currently neglected in the emerging UN framework.

Operational satellite-derived deforestation (3) and fire (15) data sets produced by INPE, and land cover information from the European Commission's Joint Research Centre (16), provide a unique opportunity to quantify the sensitivity of fires to changes in deforestation rates and land use in the Brazilian Amazon. Fire in the Brazilian Amazon is likely to follow three plausible pathways: (i) Fire incidence may decrease with reduced deforestation rates by restraining human activities that are major ignition sources (8, 13, 14, 17). (ii) It may increase even with reduced deforestation rates, both through slashing and burning of secondary forests (18) in already deforested areas that are not monitored by INPE's Program for Deforestation Assessment in the Brazilian Legal Amazonia (PRODES) (19) and through continuous enlargement of forest edges (20) and increasing area of secondary forest cover (21) that are more susceptible to fire (22). (iii) Fire incidence may decrease because of a shift from extensive (unmanaged) to intensive (managed) land-use methods, as the latter is normally not accompanied by deliberate use of fire (23).

To distinguish the first two pathways we used all available regionwide data from INPE to perform a pixel-based analysis of temporal trends in deforestation rates and fire incidence (19). For each pixel at 0.25° by 0.25° (or 774.35 km^2) spatial resolution, the annual fraction deforested for the period from 2000 to 2007 was derived by aggregating the 60-m spatial resolution pixels from INPE's PRODES annual deforestation maps (3, 19). Similarly, the annual number of fires for each 0.25° by 0.25° pixel for the period from 1998 to 2006 was derived (19) by aggregating the daily

¹Landscape and Ecosystems Dynamics Group, School of Geography, University of Exeter, Amory Building, Rennes Drive, Exeter, Devon, EX4 4RJ, UK. ²Remote Sensing Division, National Institute for Space Research, Avenida dos Astronautas, 12227-010, São José dos Campos, São Paulo, Brazil.

*These authors contributed equally to this work.

†To whom correspondence and requests for materials should be addressed. E-mail: L.aragao@exeter.ac.uk.

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1275>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1275/DC1>

This article **cites 27 articles**, 9 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/328/5983/1275#otherarticles>

This article appears in the following **subject collections**:

Ecology

<http://www.sciencemag.org/cgi/collection/ecology>

1-km spatial resolution active fire detections from the advanced very high resolution radiometer (AVHRR) aboard the National Oceanic and Atmospheric Administration NOAA-12 satellite produced by the INPES's fire monitoring system (15). For each pixel and for each data set, we computed the slope of the temporal trend regression line and the confidence level according to the two-tailed Student's t distribution (19). Finally, we analyzed the frequency distributions of pixel-based fire trends associated with positive and negative deforestation trends, in order to evaluate whether fire incidence was increasing or decreasing in each of the two deforestation categories (19).

The temporal trend analysis on the deforestation data revealed a widespread pattern of grid cells with negative slopes (54% negative against only 17% with positive slopes), indicating that deforestation rates have been decreasing in most of the Brazilian Amazon from 2000 to 2007 (Fig. 1A). Because of the overall unimodal "boom-and-bust" nature of deforestation rates (fig. S2) during the time period analyzed (24, 25), in only 4.7% of grid cells with negative slopes and 10% of grid cells with positive slopes were the trends significant at a 90% confidence level (two-tailed Student's t distribution, $t < 1.94$, $df = 8$ years $- 2$) (Fig. 1A).

Fires, on the other hand, follow a reverse pattern to deforestation with the number of grid cells with positive slopes (42%) 1.7 times those of cells with negative slopes (25%). Applying a 90% confidence level according to the two-tailed Student's t distribution ($t < 1.89$, $df = 9$ years $- 2$), our results show that in almost 30% of grid cells with an increasing fire trend [an area ~ 2.6 times larger than the U.K. (625,675 km²)], there is a statistically significant increase in fire incidence, whereas only 10% of the cells with a reducing fire trend have experienced a statistically significant decrease in fire incidence (Fig. 1B). The most drastic ($t < 2.365$, $P < 0.05$) increases in fire activity were observed in the southwest and central portions of Pará state, northwest of Mato Grosso state, north of Rondônia and Maranhão states (Fig. 1B).

By combining deforestation and fire trend results (excluding all grid cells with null slope), we were able to analyze the behavior of fire incidence in areas with increased and decreased deforestation rates (19). Most grid cells with increased deforestation rates (81%, positive regression slopes) are overlapped by grid cells with positive fire trends (Fig. 2, A and B), which confirms the expectation that fire occurrence increases with deforestation (17). Much more surprisingly, though, most of the grid cells (59%) with decreased deforestation rates (negative regression slopes) also show increased fire frequency (Fig. 2, A and B). These results were also confirmed by using the 2000 to 2006 time series for both AVHRR hot pixels and PRODES deforestation rates, as well as by using Moderate Resolution Imaging Spectroradiometer (MODIS)

fire data instead of AVHRR (19). These findings suggest that, over 31% of the study area, the pattern is more consistent with increased fire frequency despite reduced deforestation—although, over 13% of the area, the trend is consistent with increased fire with deforestation.

The trends observed here may emerge because there is a culturally widespread use of fire to manage land in Amazonia. Slash-and-burn treatment of secondary forests is a common practice in Amazonia that shifts cultivation to restore

soil properties and to reutilize the land. Secondary forest deforestation is not monitored by the PRODES system (19) (fig. S3) and, hence, is a plausible explanation of the reverse trend observed between deforestation and fire. Furthermore, reducing deforestation rates slows, but does not prevent, the continuous expansion of the cumulative area deforested. This leads to a permanent enlargement of forest edges (20) and secondary forest cover (21), which are more susceptible to fire (22). Forest landscapes in Amazonia are

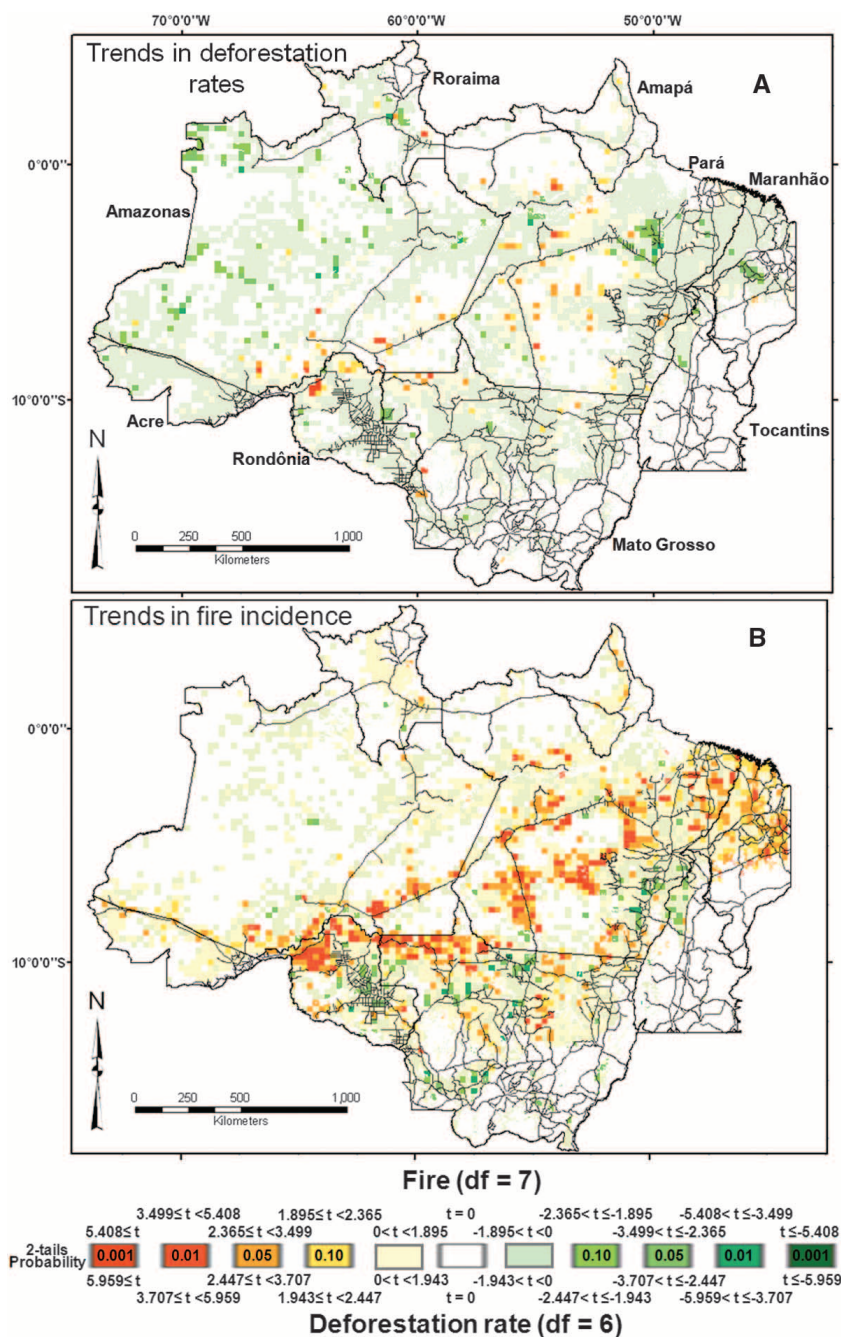


Fig. 1. Maps showing the significance of the trend regression slopes based on the two-tailed Student's t distribution for changes in (A) deforestation rates and (B) fire incidence over the time periods analyzed. From yellow to red, slopes are positive, which indicates an increase in deforestation rates or fire incidence. From light to dark green, slopes are negative, which indicates a reduction in deforestation rates or fire incidence. Different colors indicate the level of significance of the regression trend.

becoming more fragmented (20), and therefore, a growing proportion of forests is exposed to the leakage of accidental fires from adjacent farms, which could cause an increase in future fire susceptibility, fuel loading, and fire intensity in these areas (26). Furthermore, this effect is likely to catalyze a perhaps irreversible cascade of biodiversity loss by complete turnover of species composition (27), which affects the functioning

and ecology of this biome, with a consequent increase in carbon emissions to the atmosphere (28).

REDD may therefore succeed in curtailing the clearing of large original forest areas for cattle production and mechanized agriculture, which have been the foremost drivers of deforestation rates in Amazonia (29). However, although the monitored deforestation rates may be reduced,

fires and the associated carbon emissions may continue to increase because of cryptic land-use processes. So how can the patterns observed here be reversed?

To evaluate the potential of land management in regulating fire in Amazonia (our third pathway), we used a space-for-time substitution analysis (19). We first created two surfaces with 0.25° spatial resolution based on the Global Land Cover 2000 map of South America (16): (i) the proportion of total agriculture (intensive + extensive) within each grid cell, representing the current, most common land use in Amazonia (fig. S6A), and (ii) the proportion of intensive agriculture (managed agriculture) within each grid cell, as a proxy for fire-free land management (fig. S6B). We then extracted the total number of active fire detections for the year 2000 within each grid cell and compared the evolution of fire occurrence as a function of the proportion covered by each land-use category (19).

It is evident that fire incidence is higher when land starts to be cleared for intensive agriculture than for total agriculture (Fig. 3). Fire frequency associated with mechanized deforestation for commercial plantations, such as soybeans, is reported to be higher than from less intensive clearing methods (30). Nevertheless, fire incidence becomes similar in both agricultural categories when they reach between 25 and 30% of the area of the grid cell. As intensive land use begins to dominate the landscape beyond the 35% cover threshold, we observe a constant decline in fire incidence, which reaches a maximum reduction of 69% in fire occurrence when intensive agriculture covers 85% of the grid cell area (Fig. 3). Conversely, high fire incidence is maintained with the increase in total agriculture area (Fig. 3), which is characterized by a mosaic of cropland and degraded and secondary forests (19). This supports our contention that a combination of slash-and-burn of secondary forests, enlargement of forest edges, and landscape fragmentation is driving fire increase in areas with reduced deforestation rates. The continuous reduction of fuel loads with the expansion of intensive agriculture could also lead to a decrease in fire incidence; however, fire is naturally rare in Amazonia (22), and its occurrence is strongly associated with human ignition for land management (19).

Two policy-relevant conclusions can be drawn here. First, focusing on the pattern of fire incidence observed between 0% and 30% of grid cell cover, ongoing expansion of agrobusiness has the potential to drive fire increase in Amazonia and must therefore be restricted in order to prevent C emissions. Second, by analyzing fire patterns on more widely farmed areas (>35% agriculture), we find that changing land-management practices in already deforested areas, by expanding the usage of fire-free methods, can drastically reduce fire activity and associated C emissions in Amazonia. The intensification of current land management in small to medium farms could, for instance, be achieved through introduction of fire-

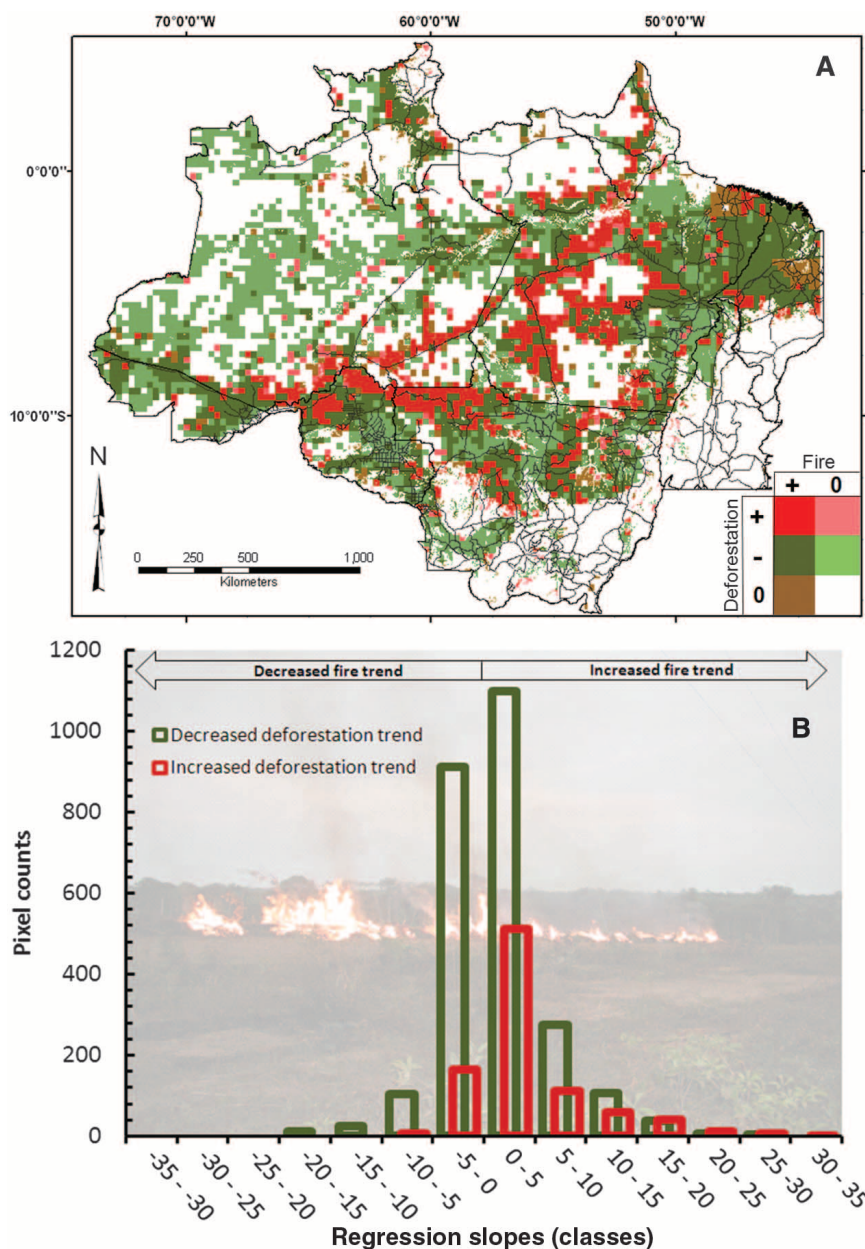
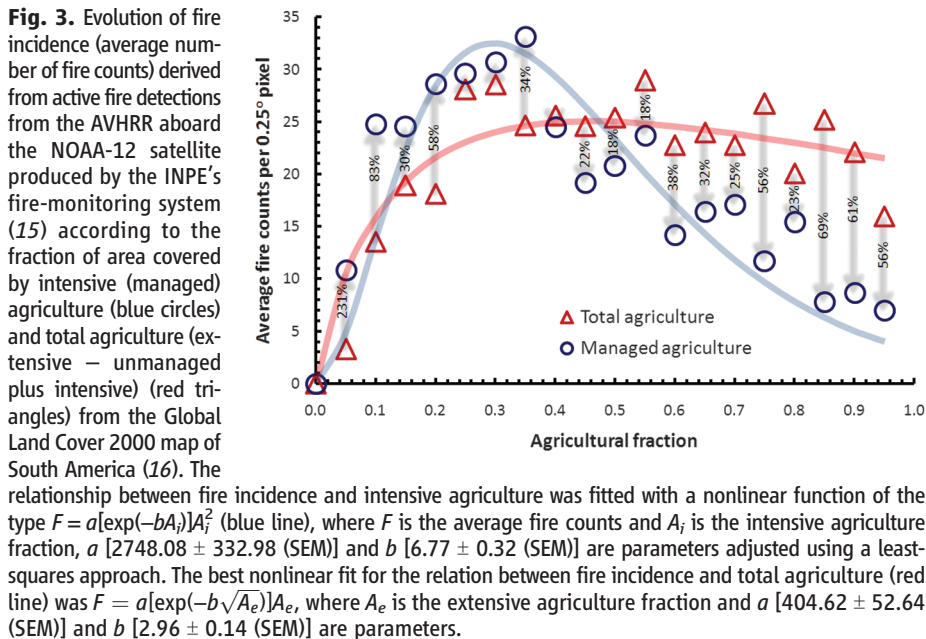


Fig. 2. (A) Pixel-based integration of the deforestation (Fig. 1A) and fire (Fig. 1B) trend surfaces obtained by using a decision rule classifier (19). The color board in the bottom left of the figure indicates the direction of the trends of each variable within the grid cell. Red cells indicate increased trend in both deforestation and fires; dark green cells indicate decreased deforestation rates and increased fire incidence. **(B)** The frequency distribution of these two major classes is shown; the red bars represent the frequency distribution of slopes for the trend regression applied for fire data over the areas with positive deforestation trend, and the dark green bars represent the frequency distribution of slopes for the trend regression applied for fire data over the areas with negative deforestation trend. Note that, in both cases, the histogram is skewed to the right, which indicates that the majority of grid cells have positive fire trends.



free methods of fallow management (31) and more diversified and sustainable agricultural and extractive practices (32) at a cooperative community level. This, however, would have consequences in further costs associated with machinery, training, and technical support to avoid leakage by emigration of farmers unable to comply with the financial demands of implementation and maintenance of fire-free management in their lands.

The success of reductions in carbon emissions by “avoiding deforestation” depends on harmonizing REDD with policies to limit fire incidence not only in the Brazilian Amazon but also in other rainforest nations in South America, Africa, and Asia. It brings to light the need for investments, in addition to the REDD finance mechanism, that aim to support “eco-friendly” land-use practices within local communities and Amazonian farmers and for monitoring systems that permit quantification of different types of forest degradation and secondary forest dynamics. Failure to tackle fire use in this region may discourage investors and donors within the REDD framework because of the risk that gains through deforestation reduction may be outweighed by carbon losses resulting from fire, as well as because of the lack of a comprehensive and reliable system for monitoring, reporting, and verifying emissions (MRV). Furthermore, fires in unmanaged forests as well as accidental fires that may not be classified as direct human-induced degradation are likely to go unreported if MRV processes mirror those on land use, land-use change, and forestry used by Annex I countries (19).

References and Notes

1. N. Stern, *The Economics of Climate Change: the Stern Review* (Cambridge Univ. Press, Cambridge, 2006); www.hm-treasury.gov.uk/independent_reviews/stern_review_economics_climate_change/stern_review_report.cfm.

2. R. E. Gullison *et al.*, *Science* **316**, 985 (2007).
3. Instituto Nacional de Pesquisas Espaciais (INPE) PRODES: Program for Deforestation Assessment in the Brazilian Legal Amazonia; www.obt.inpe.br/prodes/index.html.
4. R. S. DeFries *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14256 (2002).
5. G. R. van der Werf *et al.*, *Nat. Geosci.* **2**, 737 (2009).
6. S. S. Saatchi, R. A. Houghton, R. C. Dos Santos Alvala, J. V. Soares, Y. Yu, *Glob. Change Biol.* **13**, 816 (2007).
7. R. A. Houghton *et al.*, *Nature* **403**, 301 (2000).
8. D. C. Nepstad *et al.*, *Nature* **398**, 505 (1999).
9. O. L. Phillips *et al.*, *Science* **323**, 1344 (2009).
10. Y. Malhi *et al.*, *Science* **319**, 169 (2008).
11. J. H. Christensen *et al.*, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. D. Solomon *et al.*, Eds. (Cambridge Univ. Press, Cambridge, 2007), chap. 11.
12. Y. Malhi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20610 (2009).
13. L. E. O. C. Aragão *et al.*, *Geophys. Res. Lett.* **34**, L07701 (2007).
14. A. Alencar, D. C. Nepstad, M. C. Vera Diaz, *Earth Interact.* **10**, 1 (2006).
15. Instituto Nacional de Pesquisas Espaciais (INPE) Queimadas project; www.cptec.inpe.br/queimadas/ (2006).
16. H. D. Eva *et al.*, *Glob. Change Biol.* **10**, 731 (2004).
17. L. E. O. C. Aragão *et al.*, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 1779 (2008).
18. P. M. Fearnside, R. I. Barbosa, P. M. L. A. Graça, *For. Ecol. Manage.* **242**, 678 (2007).
19. Materials and methods are available as supporting material on Science Online.
20. E. N. Broadbent *et al.*, *Biol. Conserv.* **141**, 1745 (2008).
21. T. Neef, R. M. Lucas, J. R. Santos, E. S. Brondizio, C. C. Freitas, *Ecosystems* (N. Y.) **9**, 609 (2006).
22. M. A. Cochrane, *Nature* **421**, 913 (2003).
23. H. Eva, E. F. Lambin, *J. Biogeogr.* **27**, 765 (2000).
24. R. M. Ewers, W. F. Laurance, C. M. Souza, *Environ. Conserv.* **35**, 303 (2008).
25. A. S. Rodrigues *et al.*, *Science* **324**, 1435 (2009).
26. M. A. Cochrane *et al.*, *Science* **284**, 1832 (1999).
27. J. Barlow, C. A. Peres, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 1787 (2008).
28. D. Nepstad *et al.*, *Glob. Change Biol.* **10**, 704 (2004).
29. D. S. Alves, *Int. J. Remote Sens.* **23**, 2903 (2002).
30. D. Morton *et al.*, *Glob. Change Biol.* **14**, 2262 (2008).
31. M. Denich, P. L. G. Vlek, T. D. de Abreu Sá, K. Vielhauer, W. Lücke, *Agric. Ecosyst. Environ.* **110**, 43 (2005).
32. G. Carvalho *et al.*, *Environment* **44**, 34 (2002).
33. We thank the U.K. Natural Environment Research Council for providing funds to L.E.O.C.A. (NE/F015356/2). We acknowledge the valuable discussions with L. Anderson, T. Quine, and P. Cox during elaboration of this manuscript and thank A. Lima for providing the MODIS fire data set for the supporting analysis. We thank the Brazilian Institute for Space Research for making available their deforestation and fire data sets and the European Commission's Joint Research Centre for making available their land cover map.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5983/1275/DC1
Materials and Methods
SOM Text
Figs. S1 to S12
Table S1
References

11 January 2010; accepted 4 April 2010
10.1126/science.1186925

Meiotic Recombination Provokes Functional Activation of the p53 Regulatory Network

Wan-Jin Lu,¹ Joseph Chappo,¹ Ignasi Roig,^{2*} John M. Abrams^{1†}

The evolutionary appearance of p53 protein probably preceded its role in tumor suppression, suggesting that there may be unappreciated functions for this protein. Using genetic reporters as proxies to follow in vivo activation of the p53 network in *Drosophila*, we discovered that the process of meiotic recombination instigates programmed activation of p53 in the germ line. Specifically, double-stranded breaks in DNA generated by the topoisomerase Spo11 provoked functional p53 activity, which was prolonged in cells defective for meiotic DNA repair. This intrinsic stimulus for the p53 regulatory network is highly conserved because Spo11-dependent activation of p53 also occurs in mice. Our findings establish a physiological role for p53 in meiosis and suggest that tumor-suppressive functions may have been co-opted from primordial activities linked to recombination.

The p53 gene family mediates adaptive responses to genotoxic stress (1–3) and is broadly conserved (4, 5). It is widely

accepted that the p53 regulatory network is generally compromised in human cancers, but several lines of evidence indicate that during

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1278>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1278/DC1>

This article **cites 29 articles**, 9 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/328/5983/1278#otherarticles>

This article appears in the following **subject collections**:

Cell Biology

http://www.sciencemag.org/cgi/collection/cell_biol

evolution, animals rarely lived long enough to experience cancer (6, 7). Therefore, tumor-suppressive functions associated with p53 were probably derived from primordial activities that are poorly understood. We used *Drosophila* to investigate physiological properties of this network because, in this organism, a single p53 member exists and, like its mammalian counterparts, it coordinates stress responses and promotes genome stability (8–12).

We constructed transgenes (fig. S1) that place green fluorescent protein (GFP) under the control of an enhancer taken from sequences upstream of the *Drosophila reaper (rpr)* locus, which includes a p53 consensus binding site (13, 14). Two transgenic lines were produced and designated as “p53Rps” for p53 reporter strains. One strain produces nuclear localized GFP (p53R-GFPnls), and the other does not (p53R-GFPcyt).

Ionizing radiation (IR) leads to DNA damage and activates p53 in various model systems. Therefore, we exposed p53Rps transgenic embryos to IR and followed GFP expression by time-lapse live imaging. GFP was observed as early as 70 min after exposure to IR and was prominent at 180 min in virtually all embryos (Fig. 1A and movie S1). p53Rps expression was not observed in unirradiated embryos, nor was it observed in flies lacking *p53* or *chk2* (Fig. 1B), its upstream activating kinase (15). To confirm that damaged DNA is responsible for activation, we also observed induction of GFP in response to ultraviolet radiation and injected DNA fragments (fig. S2). Thus, the p53Rps reporters serve as authentic proxies that enable us to monitor p53 activation in live animals in real time.

We surveyed reporter activity throughout development and found little or no evidence for unstimulated expression, but, surprisingly, transient p53Rps expression was observed in germline precursors of all females. Activity was localized in regions 2a and 2b of the germarium and was absent beyond region 3 and in all egg chambers (Fig. 2). We also examined p53Rps in the germaria of *p53*^{−/−} or *chk2*^{−/−} animals and found no expression (Fig. 2B and fig. S4). Together, these observations indicate that a pulse of p53 activity occurs in oocyte precursors.

Meiotic recombination is initiated in regions 2a and 2b by Spo11 (also known as *mei-W68*), a topoisomerase that generates the DNA double-strand breaks (DSBs) needed for strand exchange (16). Therefore, we tested p53Rps activity in germaria lacking Spo11. Activation of p53 in regions 2a and 2b was absent in *spo11*^{−/−} ovaries (Fig. 2C).

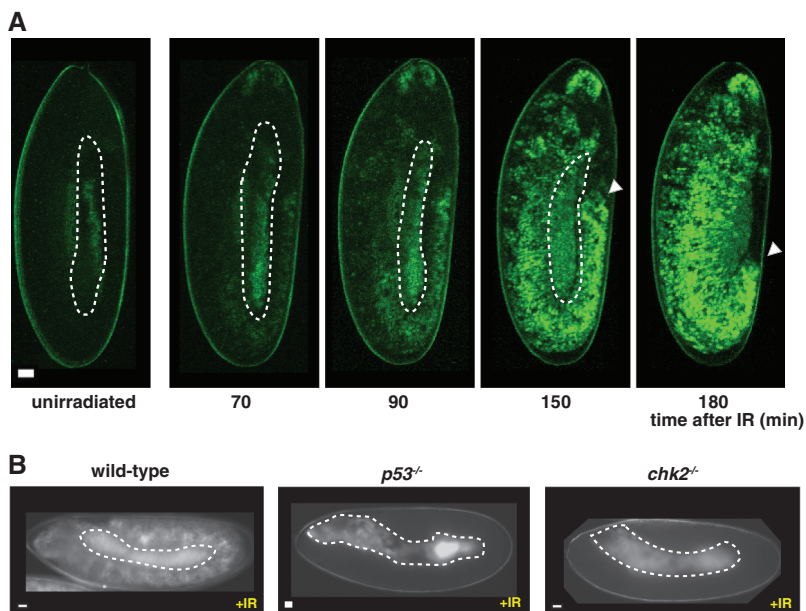


Fig. 1. Transgenic reporters as in vivo surrogates for p53 activation. **(A)** Live imaging of p53R-GFPnls expression after IR challenge at indicated time points (see also movie S1). Autofluorescent yolk material is marked with white dotted lines. Note that IR challenge did not interfere with germ-band retraction (white arrowheads). **(B)** Stimulus-induced p53R-GFPnls expression was seen in WT animals, but not in *p53*^{−/−} or *chk2*^{−/−} animals. Scale bars, 10 μ m.

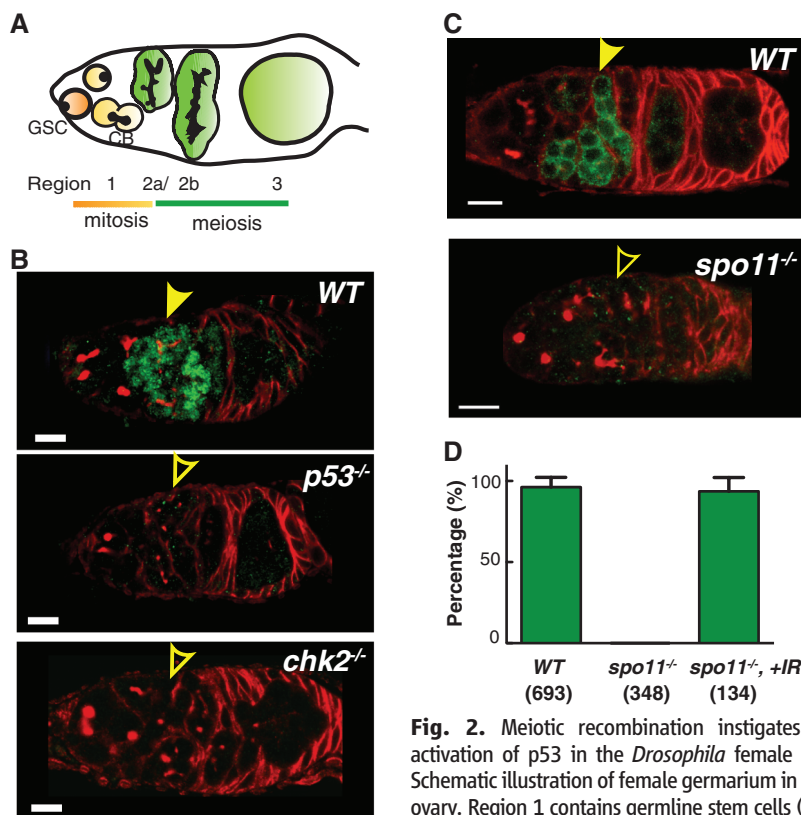


Fig. 2. Meiotic recombination instigates programmed activation of p53 in the *Drosophila* female germ line. **(A)** Schematic illustration of female germarium in the *Drosophila* ovary. Region 1 contains germline stem cells (GSC) and their progeny, cystoblasts (CB). Cysts in regions 2a and 2b initiate meiosis and further develop into egg chambers in region 3. Regions in the germarium are visualized with the use of immunostaining for *hu-li tai shao* (HTS) protein. **(B and C)** Genotypes as indicated were stained for GFP (green) and HTS (red). Animals carry p53R-GFPnls or p53R-GFPcyt in (B) or (C), respectively. Stereotyped GFP activation in region 2 is indicated by solid arrowheads; the absence of staining is indicated by open arrowheads. Scale bars, 10 μ m. **(D)** Percentage of germarium with p53R-GFPcyt expression in regions 2a and 2b. Means $\pm$ SDs (error bars) from at least three independent trials are plotted, and sample size is denoted within parentheses.

¹Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. ²Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA.

*Present address: Departamento Biología Celular, Fisiología e Inmunología, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona 08193, Spain.

†To whom correspondence should be addressed. E-mail: john.abrams@utsouthwestern.edu

Fig. 3. Functions of p53 activity during meiosis. **(A)** Meiotic recombination frequency in WT and *p53*^{-/-} flies. Recombination frequency is reduced in all p53 transallelic combinations when compared with the wild type (*Canton S* and *yw*) at all three intervals. Arrows show expected frequencies (31). **(B)** Immunostaining of GFP in p53R-GFPcyt animals that are WT or *rad54* mutants. Persisting GFP expression beyond region 2 (white dotted circle) is observed in *rad54* mutants (denoted by red arrowheads). Yellow dotted circle, region 3; yellow dotted line, egg chambers. Scale bars, 10 μm. **(C)** The incidence of p53Rps expression was quantified in region 3 and in stage two to eight egg chambers, as indicated. Error bars denote SDs from three trials. **(D)** Images of egg chambers stained with DAPI (4',6'-diamidino-2-phenylindole) from indicated genotypes. Red stars mark individual nurse cell nuclei. Scale bars, 10 μm. **(E)** Distribution of the number of nurse cell nuclei per egg chamber (nuclei/EC). Single-gene mutants show 15 nuclei/EC, but *p53*^{-/-},*rad54*^{AA/RU} ovaries exhibit a broad distribution, ranging from 9 to 40 nuclei/EC, which is restored to normal in *spo11*^{-/-},*rad54*^{AA/RU},*p53*^{-/-} animals (blue group).

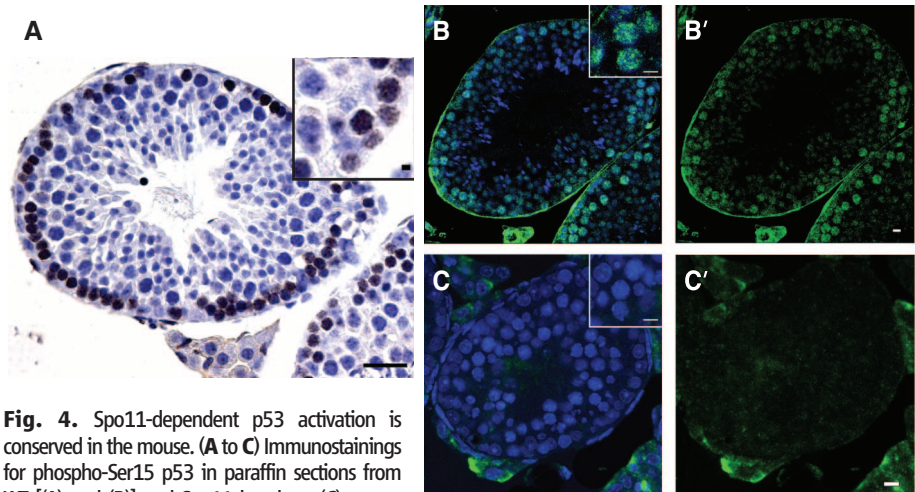
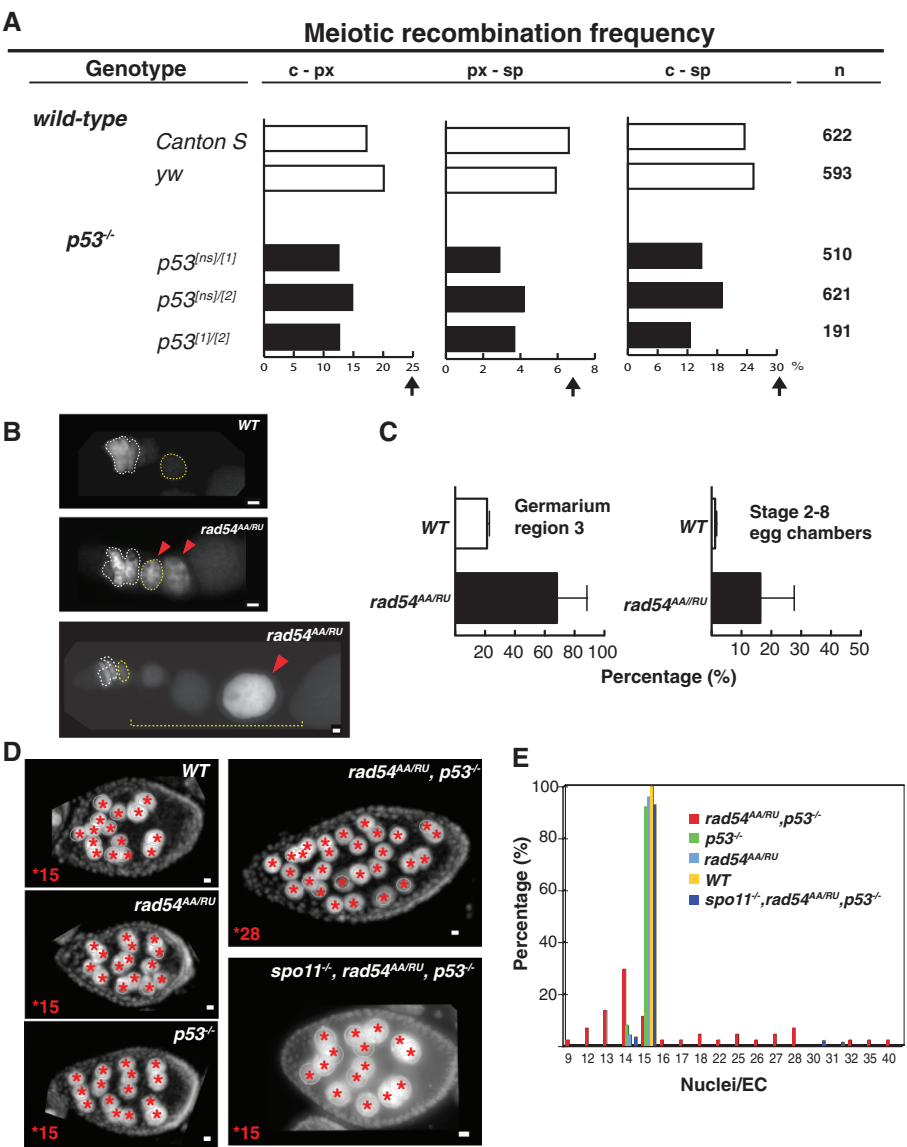


Fig. 4. Spo11-dependent p53 activation is conserved in the mouse. **(A to C)** Immunostainings for phospho-Ser15 p53 in paraffin sections from WT **(A)** and **(B)** and Spo11 knockout **(C)** mouse testes. DAB (3,3'-diaminobenzidine) **(A)**, brown and fluorescent **(B)** and **(C)**, green detection methods show that transient activation of p53 in meiotic cells is absent in Spo11-deficient testes. DNA staining (hematoxylin or Hoechst) is shown in blue. **(B'** and **C')** Signal without Hoechst counterstain. Insets in **(A)** and **(B)** show nuclear localized staining of Ser15-p53 under comparable magnifications. Scale bars, 100 μm **(A)**; 10 μm **(B)** and **(C)**, insets in **(A)**.

Furthermore, DSBs introduced by IR were sufficient to restore p53Rps activity in regions 2a and 2b of *spo11*^{-/-} germaria (Fig. 2D). Thus, DSBs formed during the initiating steps of meiotic recombination provoke activation of p53.

Flies lacking *p53* are viable and fertile. Therefore, the gene does not exert essential functions needed for gametogenesis or proper chromosome segregation (table S2). To test whether meiotic recombination frequencies were altered in *p53*^{-/-} animals, we assayed three distinct *p53* alleles in trans-allelic combinations for meiotic exchange frequencies. Reduced crossover rates, ranging from 21 to 54%, were observed across all intervals (Fig. 3A and table S1).

To understand the function of p53 during meiotic recombination, we monitored p53 activity in mutants defective for proper meiosis. Rad54 is required to properly resolve DNA crossovers, and in the germaria of *rad54* (*okra*) mutants, unrepaired DNA breaks abnormally persist (17). In such animals, we found persistent activation of

p53 beyond region 2 (Fig. 3, B and C), as indicated by p53Rps expression in region 3 of the germarium and later-staged egg chambers, where a substantial increase in the incidence of reporter activation was observed (Fig. 3C). To test whether persistent p53 activation is functionally relevant, we examined animals doubly mutated for *p53* and *rad54*. *p53^{-/-},rad54^{AA/RR}* females were sterile, despite the fact that corresponding trans-allelic combinations of the single-gene mutants were fertile (see methods for allele descriptions). Furthermore, severe oogenesis defects, which are not seen in single mutants, were evident in double mutants, including abnormal numbers of nurse cell nuclei (Fig. 3D) and shortened egg lengths (fig. S5). To test whether genetic interactions between *p53* and *rad54* were instigated by meiotic recombination, we created *spo11,p53,rad54* triple mutants. In these animals, fertility and normal nurse cell numbers were restored (Fig. 3, D and E), and defects in egg length were suppressed (fig. S5). Hence, genetic interactions between *p53* and *rad54* required the action of *spo11*. Furthermore, these results indicate that failure to properly resolve meiotic recombination can lead to sustained and functionally relevant p53 activity.

To determine whether activation of p53 by meiotic recombination is conserved, we examined mouse testes with antibodies that specifically detect phosphorylated p53 at Ser15 (18). In seminiferous tubules from wild-type (WT) animals, p53 was transiently activated in early spermatocytes (Fig. 4, A and B). On the basis of nuclear morphology (described in supplemental methods), staining appeared to peak between the leptotene and zygotene stages and disappeared after the early pachytene stage. This pattern is consistent with earlier studies on Spo11 activity (19) and p53 promoter-driven chloramphenicol acetyltransferase transgenes in this tissue (20). Staining for phospho-p53 (Ser15) was absent from testes of Spo11-deficient mice (Fig. 4, B and C), where spermatocytes survive through the leptotene and zygotene stages (21). Therefore, we can exclude cell loss as a reason for the absence of signal, and, as in flies, meiotic recombination is necessary to provoke p53 activation.

We demonstrate that a defining step in sexual reproduction, meiosis, signals a programmed burst of p53 activation. This activity is stimulated by the action of Spo11 in *Drosophila* females (recombination does not occur in males) and in mice. In flies, activation requires the kinase *chk2* [which also has roles later in oogenesis (22)] but appears to be independent of the ATM and ATR kinases (fig. S3), suggesting that alternative transducers could mediate Spo11-dependent activation of p53.

In mice (Fig. 4), potential functions of Spo11-mediated activation of p53 are indicated by altered kinetics of gametogenesis in p53-deficient mice (23, 24), as well as giant-cell degenerative syndrome in the testes of p53-deficient males (25) and implantation defects in females (26). Additional layers of complexity or redundant activities conferred by the p63 and p73 paralogs (27, 28)

could obscure conserved functions, because recombination appeared normal without p53 (29). Nevertheless, these findings raise the possibility that the act of recombination during meiosis may have been an intrinsic primordial stimulus that shaped ancestral features of the p53 regulatory network. Future studies could elucidate whether p53 directly affects crossover reactions (30) or imposes quality control on selected gametes.

References and Notes

- C. J. Brown, S. Lain, C. S. Verma, A. R. Fersht, D. P. Lane, *Nat. Rev. Cancer* **9**, 862 (2009).
- A. J. Levine, M. Oren, *Nat. Rev. Cancer* **9**, 749 (2009).
- K. H. Vousden, C. Prives, *Cell* **137**, 413 (2009).
- W.-J. Lu, J. F. Amatruda, J. M. Abrams, *Nat. Rev. Cancer* **9**, 758 (2009).
- A. M. Nedelcu, C. Tan, *Dev. Genes Evol.* **217**, 801 (2007).
- A. Aranda-Anzaldo, M. A. R. Dent, *Mech. Ageing Dev.* **128**, 293 (2007).
- M. R. Junttila, G. I. Evan, *Nat. Rev. Cancer* **9**, 821 (2009).
- F. Akdemir, A. Christich, N. Sogame, J. Chapo, J. M. Abrams, *Oncogene* **26**, 5184 (2007).
- M. H. Brodsky et al., *Mol. Cell. Biol.* **24**, 1219 (2004).
- S. Jin et al., *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7301 (2000).
- J. H. Lee et al., *FEBS Lett.* **550**, 5 (2003).
- N. Sogame, M. Kim, J. M. Abrams, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4696 (2003).
- M. H. Brodsky et al., *Cell* **101**, 103 (2000).
- Materials and methods are available as supporting material on Science Online.
- M. Peters et al., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11305 (2002).
- K. S. McKim, A. Hayashi-Hagihara, *Genes Dev.* **12**, 2932 (1998).
- M. Klovstad, U. Abdu, T. Schüpbach, *PLoS Genet.* **4**, e31 (2008).
- S. Y. Shieh, M. Ikeda, Y. Taya, C. Prives, *Cell* **91**, 325 (1997).
- S. K. Mahadevaiah et al., *Nat. Genet.* **27**, 271 (2001).
- E. Almon et al., *Dev. Biol.* **156**, 107 (1993).
- F. Baudat, K. Manova, J. P. Yuen, M. Jasin, S. Keeney, *Mol. Cell* **6**, 989 (2000).
- U. Abdu, M. Brodsky, T. Schüpbach, *Curr. Biol.* **12**, 1645 (2002).
- T. L. Beumer et al., *Cell Death Differ.* **5**, 669 (1998).
- F. Ghafari, S. Pelengaris, E. Walters, G. M. Hartshorne, *Hum. Reprod.* **24**, 1460 (2009).
- V. Rotter et al., *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9075 (1993).
- W. Hu, Z. Feng, A. K. Teresky, A. J. Levine, *Nature* **450**, 721 (2007).
- E.-K. Suh et al., *Nature* **444**, 624 (2006).
- R. Tomasini et al., *Genes Dev.* **22**, 2677 (2008).
- K. M. Gersten, C. J. Kemp, *Nat. Genet.* **17**, 378 (1997).
- T. Habu et al., *Carcinogenesis* **25**, 889 (2004).
- For more information, see FlyBase (<http://flybase.org>).
- We thank S. Keeney for helpful discussions and reading of the manuscript and A. Ali, A. D'Brot, and A. Olivo for assisting experiments. We also thank the following individuals, laboratories, or facilities for helpful discussions, access to instruments, or providing reagents: S. Barolo, M. Buszczak, D. Castrillon, O. Cleaver, K. Hamra, J. Maines, D. McKearin, T. Schüpbach, Y. Rong, J. Shelton, T. Wilkie, R. Bachoo, T. Mashimo, B. Thaden, Live Cell Imaging Facility (Univ. of Texas Southwestern), Bloomington Stock Center (Indiana Univ.), Developmental Studies Hybridoma Bank (Univ. of Iowa). This work was supported by the National Institute of General Medical Sciences (grant R01GM072124) and the National Institute on Alcohol Abuse and Alcoholism (grant R01AA017328).

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5983/1278/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 and S2
References
Movie S1

7 December 2009; accepted 28 April 2010
10.1126/science.1185640

FCHo Proteins Are Nucleators of Clathrin-Mediated Endocytosis

William Mike Henne,*† Emmanuel Boucrot,†‡ Michael Meinecke, Emma Evergren, Yvonne Vallis, Rohit Mittal, Harvey T. McMahon‡

Clathrin-mediated endocytosis, the major pathway for ligand internalization into eukaryotic cells, is thought to be initiated by the clustering of clathrin and adaptors around receptors destined for internalization. However, here we report that the membrane-sculpting F-BAR domain-containing Fer/Cip4 homology domain-only proteins 1 and 2 (FCHo1/2) were required for plasma membrane clathrin-coated vesicle (CCV) budding and marked sites of CCV formation. Changes in FCHo1/2 expression levels correlated directly with numbers of CCV budding events, ligand endocytosis, and synaptic vesicle marker recycling. FCHo1/2 proteins bound specifically to the plasma membrane and recruited the scaffold proteins eps15 and intersectin, which in turn engaged the adaptor complex AP2. The FCHo F-BAR membrane-bending activity was required, leading to the proposal that FCHo1/2 sculpt the initial bud site and recruit the clathrin machinery for CCV formation.

Clathrin-mediated endocytosis is the process by which cargo is internalized into vesicles with the aid of adaptors [such as the adaptor protein complex 2 (AP2)] and the coat-protein clathrin (1, 2). Amphiphysins and

sorting nexin 9 probably recruit the membrane scission protein dynamin to membranes of high curvature by their N-terminal BAR (Bin/Amphiphysin/Rvs) domains (3, 4). We investigated the possibility that membrane-sculpting proteins play

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1281>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/science.1188462/DC1>

This article **cites 19 articles**, 9 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/328/5983/1281#otherarticles>

This article appears in the following **subject collections**:

Cell Biology

http://www.sciencemag.org/cgi/collection/cell_biol

an early role in invagination even before AP2 and clathrin recruitment. We studied the F-BAR-containing protein family FCHO1/2 [Fer/Cip4 homology domain-only (FCHO) proteins 1 and 2], whose F-BAR homodimer module can recognize less extreme curvatures than BAR modules (5–7). FCHO1/2 are ubiquitously expressed (fig. S1, A and B) and have a twisted shape that is distinct from the F-BAR dimers of FBP17 and CIP4 (5–7). The yeast homolog Syp1 is recruited early to sites of actin-dependent endocytosis (8–10). We confirmed that FCHO1/2 are localized to clathrin-coated pits (CCPs) only on the plasma membrane (PM) (fig. S1, C to F). Furthermore, a FCHO signal defined where a CCP forms because it was detected before the visible appearance of clathrin or its PM-specific adaptor, AP2 (Fig. 1, A and C, and fig. S2, A and B). The FCHO1/2 signal decreased before the clathrin signal intensity reached its maximum, but in some rare cases the FCHO protein did not leave and defined sites where clathrin returned multiple times, thus marking endocytic “hotspots” (fig. S2C). In contrast, FBP17—another F-BAR protein implicated in clathrin-mediated endocytosis (6)—was

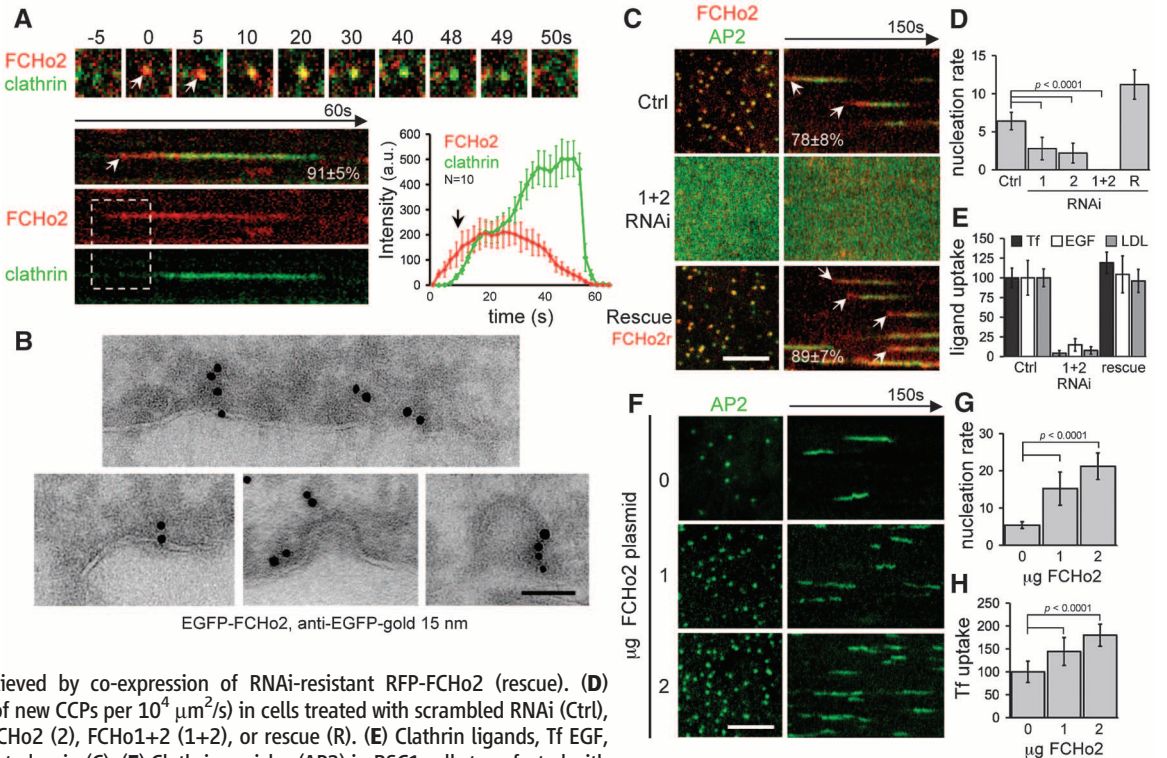
recruited at later stages to some ($3 \pm 1\%$) CCPs (fig. S2E). FCHO2 was detected with cryogenic immunoelectron microscopy (cryoimmuno-EM) at early to late stages of CCPs, which is consistent with live cell imaging kymographs (Fig. 1B). A complete loss of CCPs was observed when FCHO1+2 levels were greatly reduced by using double RNA interference (RNAi) (Fig. 1, C and D, and fig. S3, A to D) with a concomitant reduction in internalization of three known cargoes for clathrin-mediated endocytosis: transferrin (Tf), low-density lipoprotein (LDL), and epidermal growth factor (EGF) (Fig. 1E). In the absence of FCHO proteins, both AP2 and clathrin were cytosolic. These phenotypes were rescued by an RNAi-resistant form of FCHO2 (Fig. 1, C to E, and fig. S3E) (11). FCHO1/2 function was not limited to fibroblasts but was also associated with clathrin-mediated endocytosis in primary astrocytes and the recycling of synaptic vesicle markers (synaptotagmin1 and synaptophysin) after stimulated exocytosis in hippocampal neurons cultured for 4 days in vitro (fig. S4). Overexpression of FCHO1 or FCHO2 led to a dramatic increase in CCP density. This increase was not due to slowed or inhibited clathrin-coated vesicle (CCV) budding [as with epsin1 overexpression (fig. S5B) or dynamin inhibition by dynasore (12)] because CCPs were dynamic (increased nucleation rate) and functional (increased Tf uptake) during FCHO1 or 2 overexpression (Fig. 1, F to H, fig. S5, and movie S1). Because CCP numbers directly correlate with FCHO1/2 levels, FCHO proteins appear to act as CCP nucleators.

The presence of a membrane-bending protein early in CCP formation caused us to seek an explanation for how FCHO1/2 recruitment connects to clathrin recruitment. We looked for FCHO2 C-terminal AP2-μ homology domain (μHD) interactors in brain and HeLa cell extracts and found that the main interaction partners were known CCP proteins eps15, eps15R, and intersectin 1 and 2 (Fig. 2A and fig. S6, A and B). The interaction of eps15 with the μHD was direct (fig. S6C) (8) as was that with intersectin 1. The CCP localization of eps15 and intersectin was dependent on FCHO1/2 (Fig. 2B). These proteins interact directly with AP2 but not with clathrin, unlike many other CCV accessory proteins (2). Eps15 and intersectin appearance coincided with that of FCHO2 (Fig. 2C), suggesting that these proteins constitute an early module for nascent CCP assembly. FCHO1/2 are necessary for CCP formation, and yet their fluorescent intensity diminished before vesicle budding (Fig. 1A), just as dynamin intensity increased (fig. S6D). Similarly, the yeast Syp1 intensity decreased as Abp1 increased (9, 10). In purified CCVs, FCHO1/2, eps15, and intersectin levels were reduced as compared with that of total extracts (Fig. 2D and fig. S6E), which is consistent with their absence in previous mass spectrometry studies (13). Thus, FCHO1/2 initiate CCPs but are excluded from mature vesicles, with FCHO1/2 being primarily PM-associated, which is consistent with their localization on constricted CCP necks (Fig. 2E).

RNAi of AP2 leads to a marked reduction in CCP numbers (14). Thus, we tested the localization

Medical Research Council, Laboratory of Molecular Biology (MRC-LMB), Hills Road, Cambridge CB2 0QH, UK.
*Present address: Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY 14853, USA.
†These authors contributed equally to this work.
‡To whom correspondence should be addressed. E-mail: hnm@mrc-lmb.cam.ac.uk (H.T.M.); eboucrot@mrc-lmb.cam.ac.uk (E.B.)

Fig. 1. FCHO1/2 proteins are clathrin/AP2 nucleators. (A) Dynamic cell-surface localization (top) and kymograph (bottom left) of representative CCPs labeled with red fluorescent protein (RFP)–FCHO2 (FCHO2) and GFP-LCa (clathrin). FCHO2 was detected before clathrin (white arrow and bottom right graph). a.u., arbitrary units. (B) Cryoimmuno-EM-localized GFP-FCHO2 at CCPs. Scale bar, 100 nm. (C) CCVs, labeled with σ2-GFP (AP2), did not form with double RNAi of FCHO1+2 (FCHO1+2 RNAi) in which AP2 became cytosolic (contrast was enhanced so as to show the diffuse signal at the plasma membrane). Inhibition was relieved by co-expression of RNAi-resistant RFP-FCHO2 (rescue). (D) Nucleation rates (number of new CCPs per $10^4 \mu\text{m}^2/\text{s}$) in cells treated with scrambled RNAi (Ctrl), RNAi against FCHO1 (1), FCHO2 (2), FCHO1+2 (1+2), or rescue (R). (E) Clathrin ligands, Tf EGF, and LDL uptake in cells treated as in (C). (F) Clathrin vesicles (AP2) in BSC1 cells transfected with 0, 1, or 2 μg of untagged-FCHO2 for 2×10^5 cells. (G) Nucleation rate and (H) Tf uptake in cells treated as in (F). Scale bars, 5 μm (C and F) and 200 nm (B). Displayed kymographs were representative (percentage, $n = 319$ CCPs) (11).



of FCHO2 in the absence of AP2. Although AP2 puncta were largely missing, FCHO2 puncta at the PM remained and still colocalized with eps15 and intersectin (Fig. 2G and fig. S7A). The FCHO2- μ HD interactions were also AP2-independent in vitro (Fig. 2F and fig. S7B). In contrast, the localization of epsin, another membrane-sculpting molecule that binds clathrin and AP2, was dependent on the presence of AP2 (fig. S7C). An alter-

native strategy to disrupt CCP formation is the overexpression of the C terminus of AP180, which binds to clathrin with high affinity (15). Overexpression led to an accumulation of AP2 puncta, which colocalized with FCHO2, eps15, and intersectin but had no clathrin and were static (fig. S7, D to F). Thus FCHO2, eps15, and intersectin do not require AP2 or clathrin to cluster. An eps15/eps15R/intersectin1/intersectin2 quadruple knockdown af-

fected FCHO2 clustering into puncta but not its PM localization, whereas AP2 was cytosolic (Fig. 2H and fig. S8, A to E). RNAi of Dab2, a μ HD interaction partner that arrives early at CCPs (figs. S6A and S8F) and that was not enriched in CCVs (Fig. 2D), did not lead to a reduction in CCPs (fig. S8, F and G). Thus, eps15 and intersectin cluster FCHO1/2 to define nascent sites of CCP nucleation. Mutation of K797 in FCHO2 μ HD [a conserved residue that is equivalent to where AP2- β interacts with AP2- μ in pdb:2VGL (fig. S9A)] (16, 17) abolished interactions with eps15 and intersectins (Fig. 2J and fig. S9B). In the FCHO1+2 RNAi background, K797E did not rescue CCP formation and Tf uptake and was diffusely located on the PM (Fig. 2I). Thus, FCHO membrane recruitment and clustering by eps15 and intersectins initiates CCP maturation with subsequent recruitment of AP2 and clathrin, leading to coated vesicle formation.

Functionality of F-BAR domains is mediated by three distinct properties: membrane binding, dimerization, and membrane sculpting (5, 6, 18). To test the importance of each property in FCHO2 function, we designed (i) a chimera to replace the F-BAR domain with a PM-targeting PH domain, which dimerizes because of an enhanced green fluorescent protein (EGFP) tag (19); (ii) a structure-based mutant of FCHO2 (F38E+W73E), which should disrupt dimer formation; and (iii) a fluorescently tagged SGIP1, which is a close relative of FCHO1/2 and a CCP component that has a C-terminal μ HD but no F-BAR domain (20). All three proteins localized to CCPs in wild-type cells (fig. S10A) but upon depletion of endogenous FCHO1+2 did not rescue CCP formation (Fig. 3A). As expected, both the PH chimera and SGIP1 localized to the PM (sometimes in large sheets), whereas the dimer mutant remained cytosolic. Thus, the dimeric, membrane-sculpting F-BAR module is necessary for CCP formation. The region following the FCHO1/2 F-BAR domain (residues 263 to 430) is rich in positively charged amino acids and has a high homology with the N terminus of SGIP1 (9). An extended F-BAR module containing this homology region (F-BAR-x, for “extended”), showed enhanced membrane binding and tubulation in vitro (fig. S10, B and C). The F-BAR-x module co-sedimented preferentially with liposomes enriched with phosphatidylinositol 4,5-bisphosphate [PI(4,5)P₂], helping to explain why FCHO proteins are PM-targeted (Fig. 3B). Acute decrease of cellular PI(4,5)P₂ levels by the addition of 1-butanol (21) led to acute relocation of FCHO2 to the cytosol (fig. S10D), supporting the role of PI(4,5)P₂ in the targeting of FCHO1/2 to the PM. The F-BAR-x module caused extensive tubulation of PI(4,5)P₂ liposomes to high curvatures (from 130- to 18-nm tubules and many small vesicles) in a protein concentration-dependent manner (Fig. 3C). Protein density surrounding tubules sometimes exhibited striations in which the angle correlated with the degree of membrane curvature (fig. S10E and movie S2). Narrower tubules displayed more oblique angles,

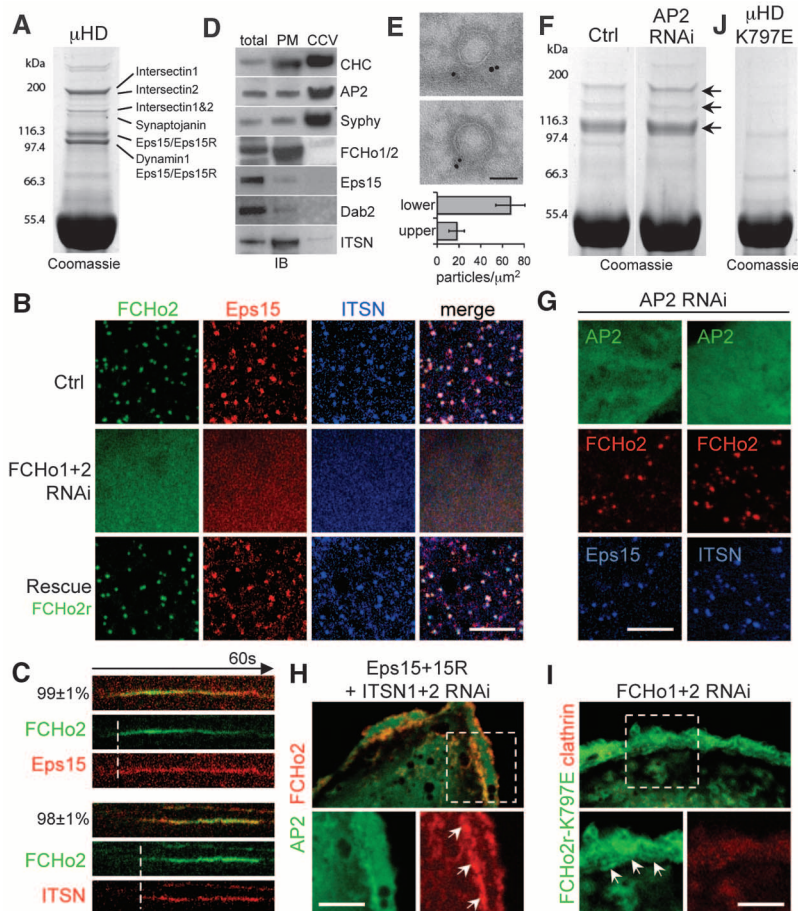
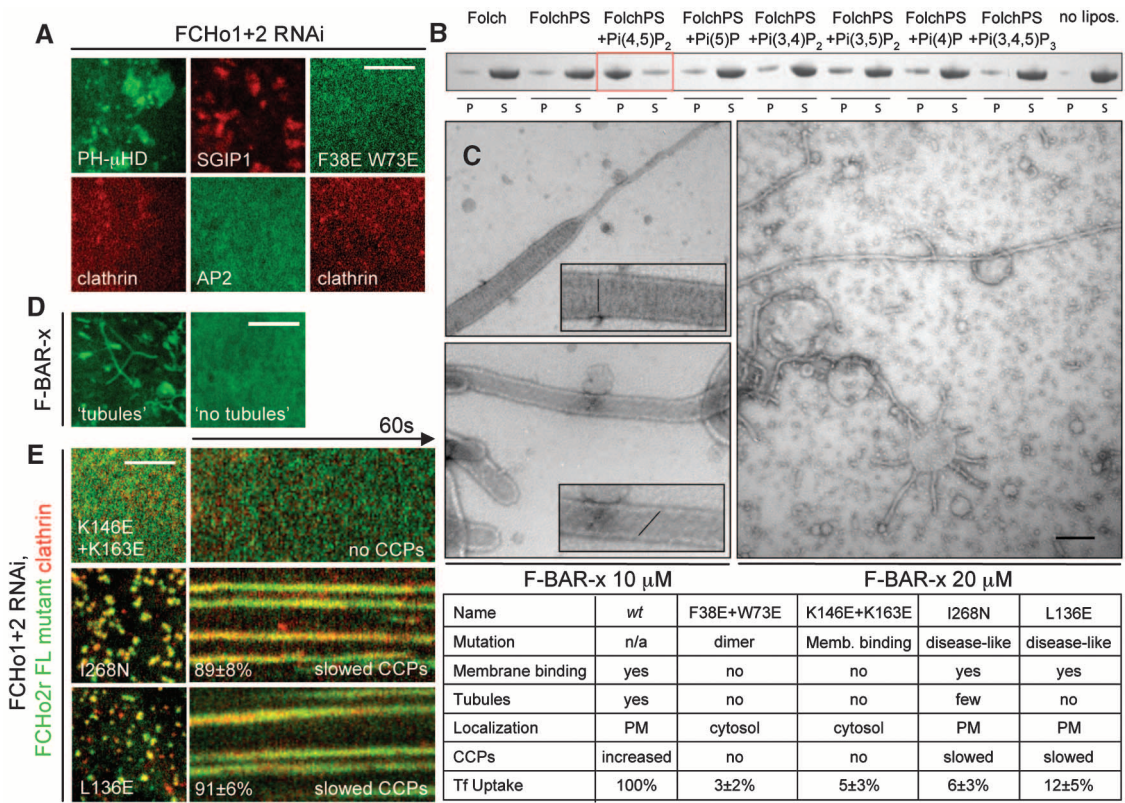


Fig. 2. FCHO2 directly binds and recruits eps15 and intersectin to initiate CCP maturation. (A) Pull-down with GST-FCHO2- μ HD and rat brain lysate. Interacting proteins were identified by means of mass spectrometry. (B) Eps15 and intersectin (ITSN) formed puncta at the PM, colocalizing with FCHO2. In double FCHO1+2 RNAi cells, eps15 and ITSN were cytosolic (contrast was enhanced so as to show the diffuse signal); co-expression of RNAi-resistant FCHO2 (FCHO2r) rescued PM-targeting (Rescue). (C) Kymograph of CCPs (percentages indicate how representative the displayed profiles are; $n = 210$ CCPs) (11) labeled with FCHO2 and Eps15 or ITSN. (D) FCHO2, eps15, and ITSN were CCV de-enriched. Clathrin (CHC), AP2, and vesicle marker synaptophysin (Syphy) displayed enrichment in CCV fractions (CCV). FCHO2 and ITSN were PM enriched. IB, immunoblot. (E) Cryoimmuno-EM of GFP-FCHO2 localized it to the CCP neck. Bar graph shows gold-particle density in the upper and lower half of constricted CCPs ($P < 0.01$). (F) Pull-downs with GST- μ HD from (left) scrambled or (right) AP2 RNAi-treated HeLa cells. Eps15 and ITSN bands were visible in both (arrows). (G) Upon AP2 depletion (μ 2 RNAi), σ 2-GFP (another AP2 subunit) was cytosolic (contrast was enhanced so as to show the diffuse signal), but Eps15 and ITSN still colocalize with FCHO2 at the PM (arrows). (H) FCHO2 and AP2 (σ 2-GFP) puncta disappeared under Eps15 + Eps15R + ITSN1 + ITSN2 quadruple RNAi. AP2 became cytosolic (diffuse signal), whereas FCHO2 remained at the PM (inset). (I) In FCHO1+2 RNAi cells, RNAi-resistant FCHO2-K797E (FCHO2r-K797E) bound to the PM (arrows) but did not cluster nor rescue CCP formation, as reported by means of RFP-LCa (clathrin). In these cells, FCHO1+2 RNAi inhibition of Tf uptake was also not rescued [$7.2 \pm 3.5\%$ of control uptake ($P < 0.0001$)]. (J) GST- μ HD K797E no longer pulled down the protein bands that are visible in (A). Green, red, and blue panels indicate GFP-, RFP-, and blue fluorescent protein (BFP)-tagged proteins, respectively. Scale bars, 5 μ m [(B), (H), (G), and (I)] and 100 nm (E).

Fig. 3. Lipid-binding and membrane-sculpting FCHo1/2 abilities are both essential for CCP formation. **(A)** Chimeric GFP-PLC-PH+FCHo2μHD (PH-μHD), a dimer interface mutant GFP-FCHo2(F38E+W73E), and RFP-SGIP1 all could not rescue CCP nucleation—monitored by following either clathrin or AP2 fluorescence—in double RNAi FCHo1+2 cells. **(B)** Lipid co-sedimentation assay of 15 μM F-BAR-x in presence of 1 mg/mL liposomes: Folch (Avanti brain lipid), FolchPS (80% Folch and 20% phosphatidylserine), or FolchPS +5% of indicated phosphatidylinositols. Liposome-bound proteins were pelleted (P) by means of ultracentrifugation, and unbound protein remained in the supernatant (S). **(C)** Folch+15%PS+5%Pi(4,5)P₂ liposomes incubated with either 10 or 20 μM F-BAR-x and spotted onto EM grids gave mainly tubules of diameters of 50 to 80 nm (10 μM) or 18 nm (at 20 μM, most were visibly twisted). (Insets) Enlarge-ments with protein density striations. **(D)** F-BAR-x-induced in vivo tubulation. Shown are representative images of “tubules” and “no tubules.” **(E)** RNAi-resistant form of full-length FCHo2 (FCHo2r) with membrane-binding mutation (K146E+K163E) remained cytosolic and could not rescue FCHo1/2



and the narrowest ones were twisted (Fig. 3C) (5, 18). This provides a mechanistic explanation for the generation of increasing curvature required for CCP budding and dynamin recruitment. To test the contribution of membrane sculpting in FCHo2 function, we mutated two conserved lysines (K146E+K165E) on the concave face of the F-BAR module as well as a conserved residue (I268N) associated with a macrophage-induced autoimmune disease (22) and its interacting residue (L136E) along the F-BAR “wing” (fig. S11, A and B). As expected, the K146E+K165E mutant displayed reduced membrane binding in vitro and relocated to the cytosol, whereas I268N and L136E mutations did not abrogate membrane binding but failed to tubulate the PM (fig. S11, C and D). When placed into full-length FCHo2, I268N and L136E induced enlarged and static aberrant CCPs and could not rescue FCHo1+2 RNAi-induced Tf uptake defect (Fig. 3E). Thus, FCHo2-mediated membrane sculpting is essential for normal CCP nucleation.

We showed that FCHo1/2 proteins nucleate CCPs and that AP2 is a later component recruiting clathrin, cargo, and accessory proteins (fig. S12). We also uncovered a role for membrane sculpting in the initiation of clathrin-mediated endocytosis. Thus, curvature generation appears to be funda-

mental to PM CCV formation from neurons to fibroblasts, and FCHo1 and -2 represent key initial proteins that ultimately control cellular nutrient uptake, receptor regulation, and synaptic vesicle retrieval.

References and Notes

1. L. M. Traub, *Nat. Rev. Mol. Cell Biol.* **10**, 583 (2009).
2. E. M. Schmid, H. T. McMahon, *Nature* **448**, 883 (2007).
3. B. J. Peter *et al.*, *Science* **303**, 495 (2004).
4. R. Lundmark, S. R. Carlsson, *J. Biol. Chem.* **278**, 46772 (2003).
5. W. M. Henne *et al.*, *Structure* **15**, 839 (2007).
6. A. Shimada *et al.*, *Cell* **129**, 761 (2007).
7. www.endocytosis.org/F-BAR_proteins/BAR-Comparisons.html.
8. A. Reider *et al.*, *EMBO J.* **28**, 3103 (2009).
9. H. E. Stimpson, C. P. Toret, A. T. Cheng, B. S. Pauly, D. G. Drubin, *Mol. Biol. Cell* **20**, 4640 (2009).
10. D. R. Boettner *et al.*, *Curr. Biol.* **19**, 1979 (2009).
11. Materials and methods are available as supporting material on Science Online.
12. E. Macia *et al.*, *Dev. Cell* **10**, 839 (2006).
13. F. Blondeau *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3833 (2004).
14. A. Motley, N. A. Bright, M. N. Seaman, M. S. Robinson, *J. Cell Biol.* **162**, 909 (2003).
15. M. G. Ford *et al.*, *Science* **291**, 1051 (2001).
16. B. M. Collins, A. J. McCoy, H. M. Kent, P. R. Evans, D. J. Owen, *Cell* **109**, 523 (2002).
17. In the mutants, other amino acids were substituted at certain locations; for example, K797E indicates that lysine at position 797 was replaced by glutamic acid.

Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

18. A. Frost *et al.*, *Cell* **132**, 807 (2008).
19. D. A. Zacharias, J. D. Violin, A. C. Newton, R. Y. Tsien, *Science* **296**, 913 (2002).
20. A. Uezu *et al.*, *J. Biol. Chem.* **282**, 26481 (2007).
21. E. Boucrot, S. Saffarian, R. Massol, T. Kirchhausen, M. Ehrlich, *Exp. Cell Res.* **312**, 4036 (2006).
22. J. Grosse *et al.*, *Blood* **107**, 3350 (2006).
23. We thank S.-Y. Peak-Chew for mass spectrometry, M. Daly for cell sorting, G. Lingley for movie design, J. E. Tyrrell for summer assistance, and Perkin Elmer for support with spinning-disk microscopy. Support was provided to H.M.M. (MRC, UK), W.M.H. (MRC-LMB Ph.D. scholarship), E.B. (Human Frontiers Science Program Organization and MRC), and M.M. and E.E. (European Molecular Biology Organization fellowships). Core funding for this work was provided by MRC.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1188462/DC1
Materials and Methods
SOM Text
Figs. S1 to S12
References
Movies S1 and S2

17 February 2010; accepted 27 April 2010
Published online 6 May 2010;
10.1126/science.1188462
Include this information when citing this paper.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
<http://www.sciencemag.org/cgi/content/full/328/5983/1285>

Supporting Online Material can be found at:
<http://www.sciencemag.org/cgi/content/full/science.1189095/DC1>

This article **cites 17 articles**, 3 of which can be accessed for free:
<http://www.sciencemag.org/cgi/content/full/328/5983/1285#otherarticles>

This article appears in the following **subject collections**:
Neuroscience
<http://www.sciencemag.org/cgi/collection/neuroscience>

The Fusogen EFF-1 Controls Sculpting of Mechanosensory Dendrites

Meital Oren-Suisa,¹ David H. Hall,² Millet Treinin,³ Gidi Shemer,^{1*†} Benjamin Podbilewicz^{1†}

The mechanisms controlling the formation and maintenance of neuronal trees are poorly understood. We examined the dynamic development of two arborized mechanoreceptor neurons (PVDs) required for reception of strong mechanical stimuli in *Caenorhabditis elegans*. The PVDs elaborated dendritic trees comprising structural units we call "menorahs." We studied how the number, structure, and function of menorahs were maintained. EFF-1, an essential protein mediating cell fusion, acted autonomously in the PVDs to trim developing menorahs. *eff-1* mutants displayed hyperbranched, disorganized menorahs. Overexpression of EFF-1 in the PVD reduced branching. Neuronal pruning appeared to involve EFF-1-dependent branch retraction and neurite-neurite autofusion. Thus, EFF-1 activities may act as a quality control mechanism during the sculpting of dendritic trees.

Morphologies of dendritic trees vary from one neuronal type to another, and the pattern of these arbors determines the manner in which a neuron processes its synaptic or sensory input. However, little is known regarding the mechanisms controlling the outgrowth and maintenance of dendritic trees (1–4). Two mechanoreceptors in *Caenorhabditis elegans* (PVDR and PVDL; right and left, respectively) are responsible for an avoidance response triggered by strong mechanical stimuli to the body (5). The complete neural system of *C. elegans* has been considered to comprise only simple-patterned neurons (6). However, recent studies show that the PVDs have a more complex morphology (7, 8).

Here, we established a genetic system to dissect the mechanisms of branch generation and plasticity of arborized neurons in *C. elegans*. To determine branching patterns, we imaged transgenic animals expressing cytoplasmic *ser-2p::GFP* (green fluorescent protein) or plasma membrane DES-2::GFP in the PVDs (table S1). The PVDs contained repetitive structural units reminiscent of multibranched candelabras or menorahs (Fig. 1A). Although the number of menorah branches varied, the menorahs appeared to develop in a stepwise manner from the L2 larva to the adult (fig. S1). The stereotypical menorah structure is likely to form a functional unit necessary for the PVDs mechanosensory activities.

Mutations in the cell fusion gene *eff-1* (9, 10) affected the pattern of PVDs arborization, resulting in disorganized and hyperbranched phenotypes (Fig. 1B). Moreover, *eff-1(ok1021)* mutant animals showed reduced sensitivity to strong mechanical stimuli (53%, $n = 106$) (11). To characterize menorah disorganization in *eff-1* mutants,

we quantified the number of processes at different degrees of the branching order (primary to senary branches; Fig. 1C and fig. S2). The frequency of secondary and tertiary branching was doubled in the *eff-1(hy21)* mutant compared with wild type. The *eff-1(hy21)* mutant had a strong branching phenotype, whereas the *oj55* mutant displayed a weaker effect, correlating with their respective epithelial fusion-defective phenotype (figs. S2

and S5). In wild-type menorahs, most sprouting of branches and bending of tertiary processes occurred at right angles to the branches of origin (Fig. 1A). In contrast, *eff-1(hy21)* mutant menorahs showed varying branching angles (Fig. 1B). We observed in the mutant a 10-fold increase in the number of branches sprouting from the secondary branch (Fig. 1D) and a 20-fold increase in the number of branches that erred and turned back 180° (Fig. 1E). These phenotypes suggest that EFF-1 sculpts and maintains right-angle nonoverlapping branches.

Cell-specific expression of *eff-1* in the PVDs (*des-2p::eff-1*) partially rescued neuronal *eff-1* phenotypes (Fig. 2C). In contrast, PVD patterning defects were not rescued by expression of *eff-1* in the neighboring epidermal tissue (fig. S3; *dpy-7p::eff-1*). In *eff-1(hy21)* mutants expressing *eff-1* in both neural and epidermal tissues (*des-2p::eff-1+dpy-7p::eff-1*), the rescue was not significantly stronger than the rescue observed with only the PVD-specific *des-2p::eff-1* (fig. S3). These rescue experiments, together with expression of EFF-1::GFP in the PVDs (fig. S4), provide evidence that *eff-1* controls branching cell autonomously. Moreover, *eff-1* overexpression in the PVDs of wild-type animals reduced branching (Fig. 2D).

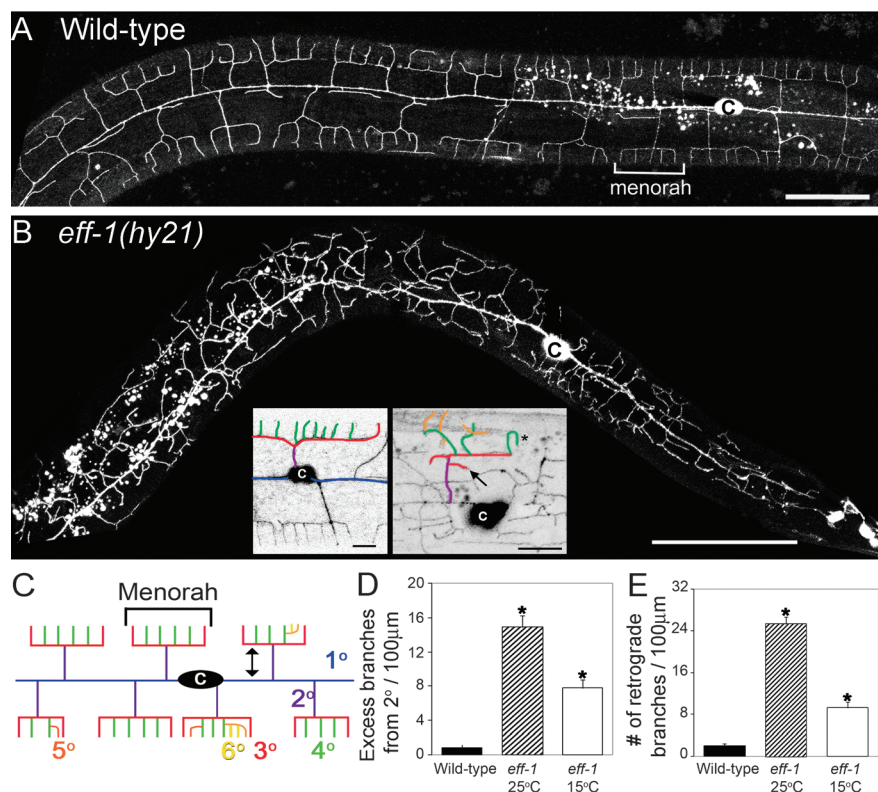


Fig. 1. PVD menorah sculpting by EFF-1. "c" denotes cell body, small droplets are autofluorescent gut granules, anterior is left, and dorsal is up. (A) Full stereotypical arborization pattern in young adult. (B) PVD in *eff-1(hy21)* showing disorganized menorahs. (Insets) Wild-type menorah (left) and *eff-1(hy21)* mutant (right) with excess branching from secondary stem (arrow) and abnormal retrograde migration of quaternary branch (asterisk). Scale bars represent 50 μm [(A) and (B)] and 10 μm [(B) insets]. (C) Menorah pattern map [primary to senary (1° to 6°) are blue, purple, red, green, orange, and yellow, respectively]. (D) Excess bifurcations from secondary stems in *eff-1(hy21)*s. (E) Increased retrograde branches in *eff-1(hy21)*s. In (D) and (E), $P < 0.0001$, two-tailed t test. Data are mean $\pm$ SE. All experiments were compared with wild type at 20°C, $n = 20$, 260 menorahs. *hy21* at 25°C, $n = 6$, 355 menorahs. *hy21* at 15°C, $n = 10$, 322 menorahs. n , number of animals analyzed. We also performed a one-way analysis of variance (ANOVA) followed by Dunnett's multiple comparison test (11).

¹Department of Biology, Technion-Israel Institute of Technology, Haifa 32000, Israel. ²Center for *C. elegans* Anatomy, Dominick P. Purpura Department of Neuroscience, Albert Einstein College of Medicine, New York, NY 10461, USA. ³Department of Physiology, Hebrew University-Hadassah Medical School, Jerusalem 91120, Israel.

*Present address: Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. †To whom correspondence should be addressed. E-mail: podbilew@tx.technion.ac.il (B.P.); bishemer@email.unc.edu (G.S.)

Fig. 2. EFF-1 autonomously rescues PVD arborization and restricts branching. (A) *eff-1(hy21)* at 25°C had twice as many secondary and tertiary branches as the wild type. Scale bars represent 10 μ m. (B) Menorahs in wild-type animals. (C) Expression of EFF-1 (*des-2p::EFF-1*) in the PVD in *eff-1(hy21)* mutants at 25°C partially rescued the menorah pattern and reduced the number of branches. (D) Overexpression of *des-2p::EFF-1* in wild-type animals caused a menorah gradient starting from the PVD cell body until dendrites disappeared toward the posterior (arrows). (E) Apparent EFF-1 dosage-sensitive reduction in branching. * $P < 0.0001$, ** $P < 0.001$, *** $P < 0.05$, two-tailed t test. We also performed ANOVA (11). Data are mean $\pm$ SE. *eff-1(hy21)*, $n = 6$; number of wild-type animals, $n = 6$; *eff-1(hy21); des-2p::EFF-1*, 25°C, $n = 5$; *des-2p::EFF-1*, $n = 6$. (F) Quantification of the number of secondary to senary branches showed that secondary and tertiary branches were doubled in the mutant compared with wild type and that secondary to quinary branches were reduced when EFF-1 was expressed ectopically in the PVD. (G) A model for the maintenance of PVD branching in an *eff-1* dosage-dependent manner. Low levels of *eff-1* (A) increased the number of PVD branches, and elevated levels of *eff-1* reduced branching (D).

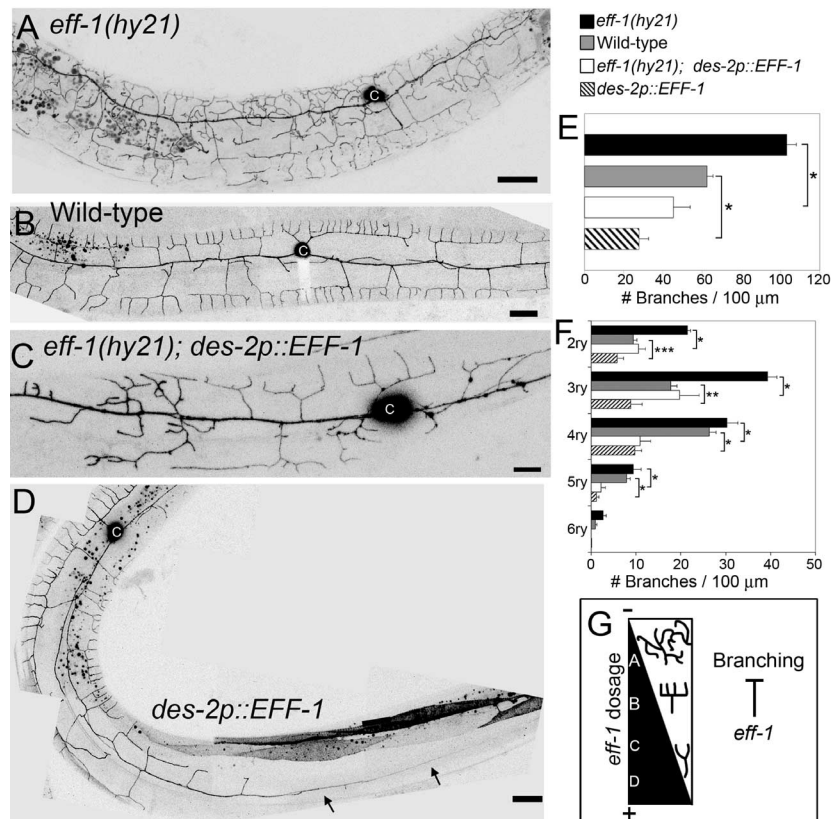
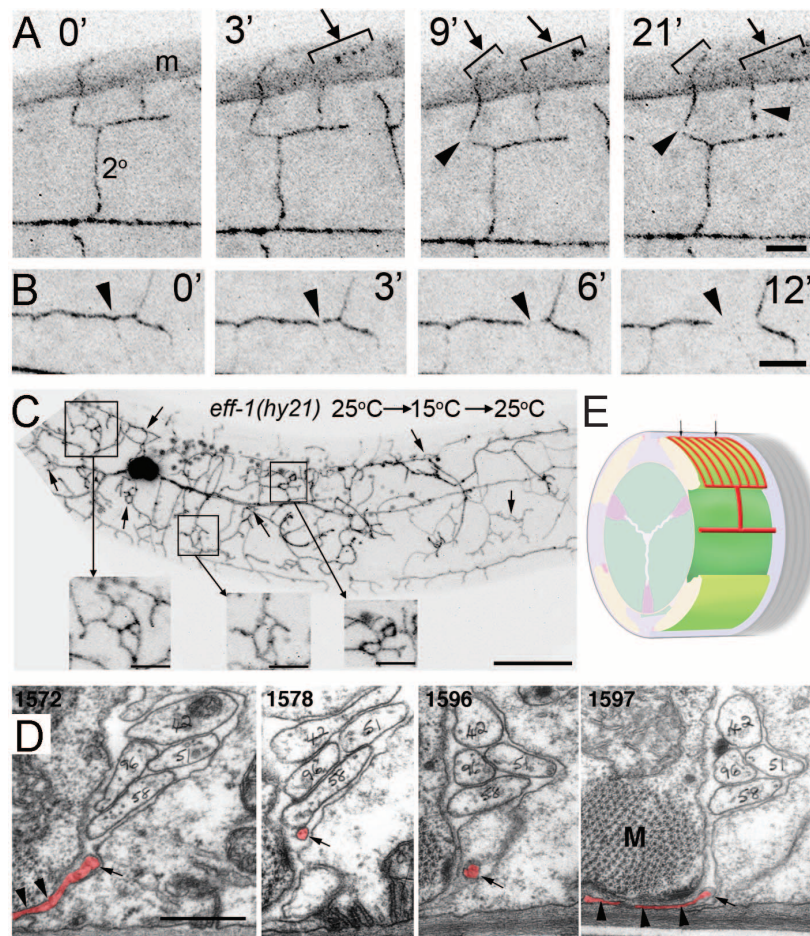


Fig. 3. Dendrite autofusion and fission trim menorahs. Scale bars represent 5 μ m, (A) and (B); 20 μ m and insets 5 μ m, (C); and 0.5 μ m, (D). (A) Trimming of dendrites that break off (arrows) and fission of branches in *des-2p::EFF-1* animal (arrowheads). m, body wall muscle. (B) Fission of a branch in *des-2p::EFF-1* animal (arrowhead). (C) *eff-1(hy21ts)* animal grown at 25°C downshifted to 15°C for 4 hours and then upshifted to 25°C for 2 hours. Loops can be seen (arrows and insets). (D) Four TEM prints from wild-type animal "N2U" (6) showing the route of a PVD fusion process at the edge of the ventral nerve cord. Prints 1572 and 1597 show quaternary dendrites (arrowheads, red pseudocolor) passing between a thin layer of hypodermis and body wall muscle (M). PVD branch emerged from one quaternary dendrite at 1572, joined the minor fascicle of the ventral cord between 1578 and 1592, and ran ventralward to autofuse with another quaternary dendrite at 1596 and 1597 (arrows). Whereas this fusion process of PVD spans about 3 μ m along the anterior-posterior axis, the marked neuron processes (42, 51, 58, and 96) were traced for thousands of serial sections back to their cell bodies. Dorsal is down (fig. S9; www.wormimage.org). (E) Cartoon representing one menorah (red) with eight terminal dendrites autofused along the dorsal midline (arrows).



The remaining branches were organized in a gradient starting from the cell body toward the head and tail, where no branches could be observed. Thus, *eff-1* may play a role in mechanosensory neurons restricting branching in a dosage-dependent manner to produce dendrite simplification (Fig. 2G).

The absence of excess branching in wild-type animals may reflect a situation where only the appropriate branches initiate outgrowth. Alternatively, an excess of branches may be generated and extended but at some later point undergo retraction, pruning, or fusion to repair branching errors. To determine how EFF-1 restricts branching, we followed

the behavior of branches by using three-dimensional (3D) live-imaging confocal microscopy in wild-type animals overexpressing *des-2p::eff-1*. The tips of tertiary to senary menorah branches first contacted the muscle cells, and then the dendrites detached from the menorahs by fission events (Fig. 3A; arrowheads) causing spontaneous dendrite break-off (Fig. 3B; arrowheads). Thus, following dynamic outgrowth, fission events eliminate extra branches.

To further analyze *eff-1*-dependent remodeling of menorahs, we grew the *eff-1(hy21ts)* worms at the restrictive temperature and shifted them to the permissive temperature in the early L4 stage.

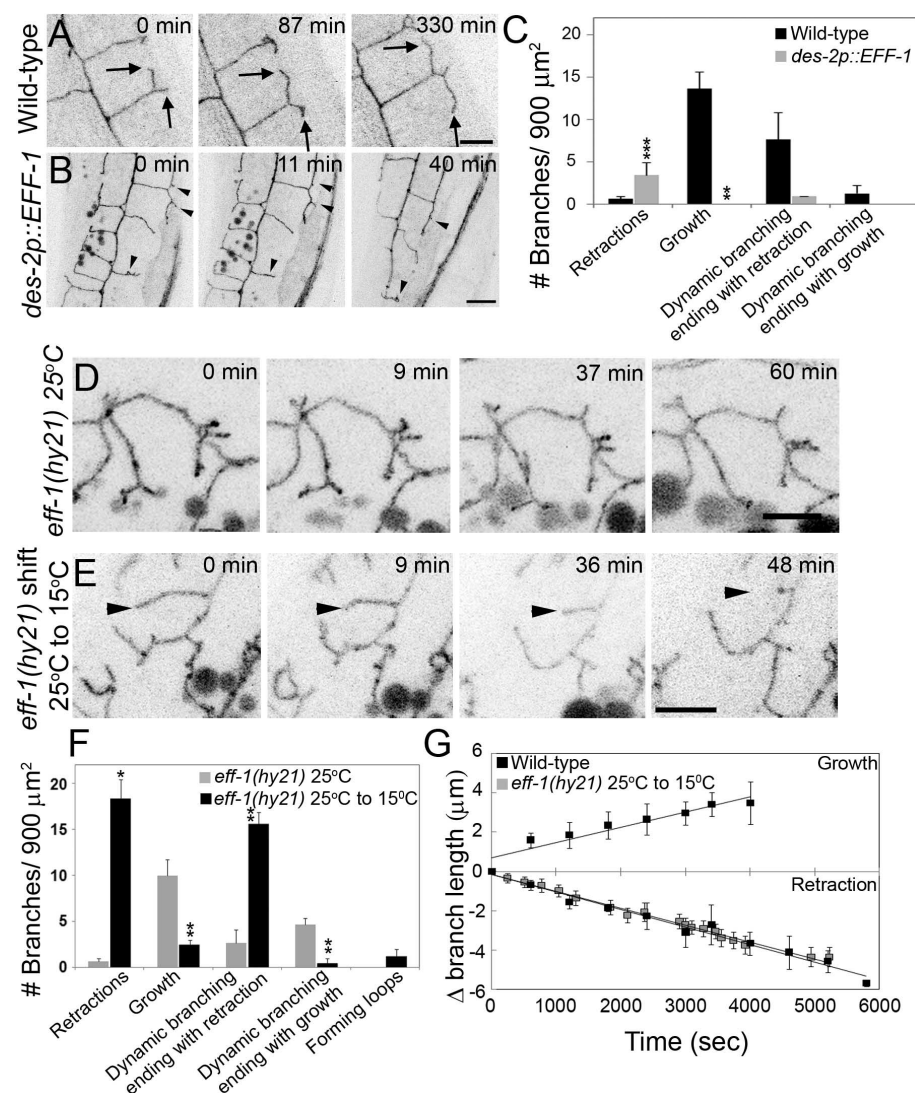


Fig. 4. EFF-1-dependent retraction of branches. (A, B, D, and E) Time-lapse confocal projections of L4 and young-adult animals. (A) Growth of a wild-type menorah (movie S3). Growing tertiary branches (arrows). (B) Retraction of branches in *des-2p::EFF-1* animal (arrowheads). (C) Number of branches growing and retracting in *des-2p::EFF-1* and wild type. Branches first showed dynamic movements but eventually either shortened or lengthened (dynamic branching ending with growth or retraction). ** $P < 0.01$, *** $P < 0.05$, two-tailed t test. Data shown as mean $\pm$ SE. Wild type, $n = 3$; *des-2p::EFF-1*, $n = 2$. (D) L4 *eff-1(hy21ts)* at 25°C, branches were static. (E) Retracting branch in L4 *eff-1(hy21ts)* grown at 25°C and shifted to 15°C for 4 hours (arrowhead; movie S1). (F) Number of branches showing growth and retraction in *eff-1(hy21)* grown at 25°C and downshifted as in (E). * $P < 0.001$, ** $P < 0.01$, two-tailed t test. Data are mean $\pm$ SE. *eff-1(hy21)* 25°C, $n = 3$; *eff-1(hy21)* 25°C to 15°C, $n = 4$ animals. (G) Wild-type branches grew (0.8 nm/s; $N = 5$ branches) and retracted (0.9 nm/s; $N = 3$ branches). In *eff-1(hy21)* temperature-shifted animals, branches retracted (0.9 nm/s; $N = 10$ branches). Data are mean $\pm$ SE. Scale bars represent 5 μm .

The defective arborization pattern was static at the nonpermissive temperature. Four hours after downshifting to the permissive temperature, we observed branches that met, touched, and retracted. Some branches appeared to be stably connected, forming loops of different sizes and shapes (43 stable loops, $n = 18$; fig. S6). Because EFF-1 is a fusogen, we hypothesized that transient meeting and attachment of branches expressing EFF-1 may have resulted in loop formation because of interbranch fusion. To distinguish between interbranch fasciculation and fusion, we analyzed the 3D structure of the loops by generating confocal z stacks, projections, and rotations. The loops were stably connected over time (55 loops, $n = 18$ animals; movie S1). Next, we imaged DES-2::GFP in loops and found that fluorescent particles moved freely through them. To test whether the loops resulted from *eff-1* activity, we incubated temperature-sensitive mutants at the restrictive temperature to generate hyperbranching, then we shifted them to the permissive temperature to induce expression of active EFF-1. After generating loops for 3.5 hours, we upshifted the worms back to the restrictive temperature to stabilize the loops (fig. S8). We obtained a two- to threefold increase in the number of loops (119 stable loops in $n = 16$ animals; Fig. 3C). We also observed symmetric fluorescence recovery of photobleached areas within the loops (fig. S7 and movie S2), indicating that the loops are continuous. Thus, *eff-1* induces loop formation by neurite autofusion and/or fasciculation.

To further demonstrate that neurites can fuse with each other, we turned to higher resolution images of menorah branches and membranes. By using archival serial section transmission electron micrographs [TEMs (6)], we identified arborizations derived from the PVDs. In transverse sections of an adult hermaphrodite (N2U), 30- to 80-nm-diameter branches sandwiched between the body wall muscles and the hypodermis were observed (Fig. 3D, arrowheads). We reconstructed parallel neurites corresponding to the quaternary branches of a menorah from serial sections (Fig. 3E and figs. S9 to S11). In 10 examples of PVDs and FLPL/R (highly arborized neurons anterior to the PVDs), we found 2 to 12 branches fusing at the midline and forming fused longitudinal processes (Fig. 3, D and E, arrows). In no cases did we observe distal dendrites fasciculate; instead these fused if they reached one another. Thus, neurite-neurite autofusion plays a role in PVD and FLP arborization.

In addition to loop formation, dendrites were highly dynamic, erring, growing, and retracting in wild-type worms (Fig. 4A and movie S3). PVDs overexpressing *eff-1* showed similar retractions but, unlike in the wild type, showed limited branch growth (Fig. 4, B and C). Similarly, when *eff-1(hy21ts)* animals were downshifted from the restrictive to the permissive temperature, we observed a 20-fold increase in the number of retracting branches and a 5-fold reduction in branch growth compared with *eff-1* mutants at 25°C (Fig. 4, D to G, and movie S1). In contrast, upshifting *eff-1(hy21ts)* worms resulted in excess neurite growth

(Fig. 4F). Thus, we propose a model in which EFF-1 autonomously induces retraction of branches to simplify menorahs.

EFF-1 is both a sculptor of epithelial organs by cell fusion (10) and a menorah sculptor by controlling dendrite bending, retraction, and fusion. The activities of this fusogen may be due to its ability to induce membrane curvature, a process that is thought to constitute a major driving force in membrane fusion and fission (12–14). Proteins capable of bending membranes, such as atlastins (15, 16) and dynamins (17), can induce tubulation, fusion, and fission (12, 13). Three mechanistic principles may form and maintain branched tubes in the cytoplasm and in extracellular branched filopodia or neuronal arbors such as menorahs: first, assembly of specialized proteins on membranes; second, membranous tube formation involving growth and bifurcation of tubes; and third, membrane bending followed by membrane fusion and fission restricts excessive branching. How can EFF-1 control mechanistically different processes such as dendrite fusion and retraction? Different isoforms and interactions may account

for diverse activities. For example, trans interactions between EFF-1 on dendrites will cause autofusion, whereas assembly of large EFF-1 complexes on dendrites may induce actin-mediated retraction.

Note added in proof: After we submitted this report, Ghosh-Roy *et al.* (18) showed that axotomized PLM sensory neurons fail to re-connect in *eff-1* mutants.

References and Notes

1. E. K. Scott, L. Luo, *Nat. Neurosci.* **4**, 359 (2001).
2. W. B. Grueber, Y. N. Jan, *Curr. Opin. Neurobiol.* **14**, 74 (2004).
3. D. M. Tobin, C. I. Bargmann, *J. Neurobiol.* **61**, 161 (2004).
4. C. Jinghong, T. Guoxi, *Brain Res.* **846**, 243 (1999).
5. J. C. Way, M. Chalfie, *Genes Dev.* **3**, (12A), 1823 (1989).
6. J. G. White, E. Southgate, J. N. Thomson, S. Brenner, *Philos. Trans. R. Soc. London Ser. B* **314**, 1 (1986).
7. E. L. Tsalik *et al.*, *Dev. Biol.* **263**, 81 (2003).
8. S. Halevi *et al.*, *EMBO J.* **21**, 1012 (2002).
9. W. A. Mohler *et al.*, *Dev. Cell* **2**, 355 (2002).
10. B. Podbilewicz *et al.*, *Dev. Cell* **11**, 471 (2006).
11. Materials and methods are available as supporting material on Science Online.
12. S. Martens, H. T. McMahon, *Nat. Rev. Mol. Cell Biol.* **9**, 543 (2008).
13. L. V. Chernomordik, M. M. Kozlov, *Annu. Rev. Biochem.* **72**, 175 (2003).

14. A. Sapir *et al.*, *Dev. Cell* **12**, 683 (2007).
15. G. Orso *et al.*, *Nature* **460**, 978 (2009).
16. J. Hu *et al.*, *Cell* **138**, 549 (2009).
17. P. V. Bashkurov *et al.*, *Cell* **135**, 1276 (2008).
18. A. Ghosh-Roy, Z. Wu, A. Goncharov, Y. Jin, A. D. Chisholm, *J. Neurosci.* **30**, 3175 (2010).
19. We thank A. Fire for vectors; N. Assaf and T. Gattegno for *eff-1* alleles; E. Shmukler for preliminary phenotypic characterization of the PVD in *eff-1(hy21ts)*; T. Stiernagle and the CGC for nematode strains; J. White and J. Hodgkin for donation of the MRC/LMB archival prints to DHH; C. Crocker for TEM cartoon; and O. Avinoam, D. Cassel, M. Kozlov, A. Sapir, and I. Yanai for critically reading the manuscript. Supported by grants from the FIRST program of the Israel Science Foundation (ISF 1542/07 to B.P.), ISF 493/01-16.6 to M.T., and NIH RR12596 (WormImage Web site) to D.H.H.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1189095/DC1
Materials and Methods

Figs. S1 to S11

Table S1

References

Movies S1 to S3

4 March 2010; accepted 26 April 2010

Published online 6 May 2010;

10.1126/science.1189095

Include this information when citing this paper.

Induction of Fear Extinction with Hippocampal-Infralimbic BDNF

Jamie Peters,¹ Laura M. Dieppa-Perea,¹ Loyda M. Melendez,² Gregory J. Quirk^{1*}

The extinction of conditioned fear memories requires plasticity in the infralimbic medial prefrontal cortex (IL mPFC), but little is known about the molecular mechanisms involved. Brain-derived neurotrophic factor (BDNF) is a key mediator of synaptic plasticity in multiple brain areas. In rats subjected to auditory fear conditioning, BDNF infused into the IL mPFC reduced conditioned fear for up to 48 hours, even in the absence of extinction training, which suggests that BDNF substituted for extinction. Similar to extinction, BDNF-induced reduction in fear required *N*-methyl-D-aspartate receptors and did not erase the original fear memory. Rats failing to learn extinction showed reduced BDNF in hippocampal inputs to the IL mPFC, and augmenting BDNF in this pathway prevented extinction failure. Hence, boosting BDNF activity in hippocampal-infralimbic circuits may ameliorate disorders of learned fear.

Extinction of conditioned fear forms a new memory in the infralimbic medial prefrontal cortex (IL mPFC) that is critical for the retrieval of extinction (1, 2). IL single-unit responses correlate with the successful retrieval of such extinction memories (3), and IL stimulation strengthens these memories (3). Consolidation of extinction requires plasticity within the IL mPFC, which in turn depends on *N*-methyl-D-aspartate (NMDA) receptors, mitogen-activated protein kinase, and protein synthesis (2, 4). Understanding the molecular mechanisms that support this extinction-related plasticity could lead to pharmacological approaches for enhancing extinction memory, which might facilitate the treatment of anxiety disorders.

Epigenetic regulation within the IL mPFC of the gene encoding BDNF correlates with fear extinction (5). Because BDNF is a major molecular mediator of memory consolidation (6), we hypothesized that BDNF is responsible for consolidating extinction memory within the IL mPFC. If true, it should be possible to enhance extinction via direct application of BDNF to the IL mPFC. Accordingly, rats were subjected to auditory fear conditioning and, the following day, received bilateral IL mPFC infusion of human recombinant BDNF protein (0.75 μ g per side) 60 min before extinction training. Conditioned freezing in BDNF-treated rats was significantly reduced relative to saline-infused rats (main effect of drug $F_{1,14} = 28.359$, $P < 0.001$, Fig. 1A; for suppression of food seeking, see fig. S1). This effect persisted in an extinction test the following day (day 3, main effect of drug $F_{1,14} = 11.029$, $P = 0.005$, Fig. 1A), which indicated that BDNF strengthened extinction memory.

Freezing was significantly reduced in BDNF rats from the first extinction trial [$t(14) = 3.335$,

$P = 0.005$], which suggested that BDNF reduced fear independent of extinction training. We therefore repeated the previous experiment but omitted extinction training from day 2. Conditioned rats were infused with BDNF or saline and returned to their home cages. The following day, freezing was again reduced in BDNF-treated rats from the first trial [$t(10) = 4.476$, $P = 0.001$, Fig. 1B] and throughout the extinction session (main effect of drug $F_{1,10} = 27.220$, $P < 0.001$). Although the effect of BDNF on fear did not require extinction training, it did require conditioning, because BDNF infused 1 day before conditioning did not significantly reduce freezing (Fig. 1C). BDNF infusions did not alter locomotion, anxiety, or motivation to seek food reward (fig. S2, A to C). The lack of effect on conditioning and open-field anxiety suggests that BDNF infusions did not decrease amygdala activity nonspecifically. Nor could BDNF's effects be attributed to potentiation of latent inhibition, because removing habituation trials did not prevent the effect (fig. S2D).

There are two interpretations for these results. BDNF could inhibit fear expression (similar to extinction), or it could have degraded the original fear memory. To distinguish between these possibilities, we determined the extent to which freezing could be reinstated after unsignaled footshocks, which can reveal the underlying fear memory (7). One day after infusions, rats were given extinction training followed by two unsignaled shocks. Replicating our previous experiment, BDNF rats showed reduced fear throughout the extinction session (main effect of drug $F_{1,21} = 7.337$, $P = 0.013$, Fig. 2A). On day 4, however, both saline- and BDNF-treated rats froze equivalently to the tone (78% and 80%, respectively; Fig. 2A), indicating that BDNF left the original fear memory intact. The return of freezing on day 4 was not due

¹Department of Psychiatry and Department of Anatomy and Neurobiology, University of Puerto Rico School of Medicine, San Juan, PR 00936, USA. ²Department of Microbiology, University of Puerto Rico School of Medicine, San Juan, PR 00936, USA.

*To whom correspondence should be addressed. E-mail: gjquirk@yahoo.com

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1288>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1288/DC1>

This article **cites 27 articles**, 13 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/328/5983/1288#otherarticles>

This article appears in the following **subject collections**:

Neuroscience

<http://www.sciencemag.org/cgi/collection/neuroscience>

to BDNF “wearing off” (fig. S3A) or contextual conditioning (fig. S3B).

One hallmark of extinction memory is its dependence on NMDA receptors (4, 8, 9). For example, systemic administration of the NMDA receptor antagonist 3(2-carboxypiperazin-4-yl)-propyl-1-phosphonic acid (CPP) prevents long-term extinction memory (10). The BDNF receptor TrkB interacts with the NMDA receptor in vivo (11), and BDNF enhances NMDA currents in vitro (12). It is possible, therefore, that IL BDNF mediates its extinction-like effects through NMDA receptors. To test this, we conditioned rats as previously on day 1. On day 2, in the absence of training, rats received one of the following treatment combinations: (i) saline injection (intraperitoneally) + saline infusion into IL (SAL + SAL), (ii) saline injection + BDNF infusion (SAL + BDNF), or (iii) CPP injection + BDNF infusion (CPP + BDNF). On day 3, all rats were returned to the chambers for a single-tone test. As before, SAL + BDNF rats showed significantly reduced fear relative to SAL + SAL rats (main effect of drug $F_{2,25} = 4.597$, $P = 0.020$, post hoc $P =$

0.046; Fig. 2B). However, CPP + BDNF rats were indistinguishable from SAL + SAL rats in their freezing level (post hoc $P = 0.828$; Fig. 2B), which demonstrated that NMDA receptors are necessary for BDNF-induced reductions in fear.

Does extinction depend on endogenous BDNF levels in the IL mPFC or its inputs? We addressed this question by capitalizing on the fact that there can be considerable variability in extinction memory across rats (8, 13). Rats were conditioned and extinguished on days 1 and 2, respectively, as above. We then selected two subgroups on the basis of their ability to successfully recall extinction on day 3. “Extinction Failure” and “Extinction Success” rats had freezing values in the top or bottom 44%, respectively (i.e., the middle 12% was excluded). These two subgroups differed significantly on test

day [$t(10) = 4.728$, $P = 0.001$] but showed no significant differences during conditioning or extinction training (Fig. 3A). Normal extinction training followed by poor retrieval of extinction is consistent with impaired infralimbic function (1, 2).

For each subgroup, brain tissue from the mPFC, amygdala, and hippocampus was dissected 24 hours after the extinction test to determine BDNF levels. The amygdala and hippocampus were chosen as putative BDNF-containing inputs that might be important for supplying BDNF to the IL mPFC to facilitate extinction recall (14–16). Indeed, hippocampal CA1 neurons produce BDNF (16, 17) and project to the IL mPFC (14). BDNF protein levels in the Success group were elevated relative to the Failure group in the hippocampus [$t(9) = 4.370$, $P = 0.002$], but not the mPFC or

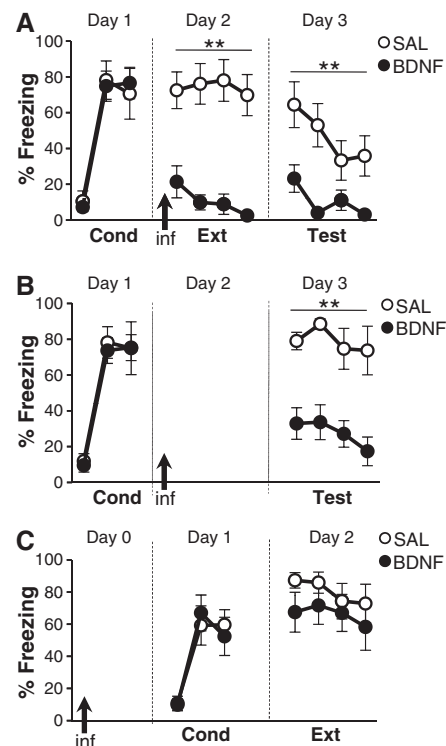


Fig. 1. BDNF infused into the infralimbic cortex substitutes for behavioral extinction. (A) Rats' freezing levels in response to tones that were paired with footshocks (Cond) or given alone during extinction (Ext) and Test sessions. BDNF infusions into the IL mPFC before extinction (arrow) reduced freezing on days 2 and 3 relative to saline-infused (SAL) controls ($n = 8$ per group). (B) A similar effect was observed when BDNF was infused in the absence of training on day 2 (SAL, $n = 5$; BDNF, $n = 7$). (C) Infusing BDNF 24 hours before conditioning had no effect (SAL, $n = 9$; BDNF, $n = 7$). Trials are shown in blocks of two. $**P < 0.01$, repeated-measures analysis of variance (ANOVA). Error bars represent SEM.

Fig. 2. Similar to extinction, the BDNF effect does not degrade the original fear memory and requires NMDA receptors. (A) Conditioned rats received BDNF or saline infusions into the IL mPFC on day 2 (SAL, $n = 12$; BDNF, $n = 11$). On day 3, both groups were extinguished, followed by two shocks, resulting in a complete return of freezing in the BDNF group. (B) IL infusion of BDNF was combined with a systemic injection of the NMDA antagonist CPP (CPP + BDNF, $n = 8$). Controls were infused with BDNF and given a saline injection (SAL + BDNF, $n = 10$) or were both infused and injected with saline (SAL + SAL, $n = 10$). On day 3, all groups underwent a single-tone extinction test. $*P < 0.05$, two-way repeated-measures ANOVA, main effect of drug; $*P < 0.05$, Student's t test, SAL + SAL compared to SAL + BDNF.

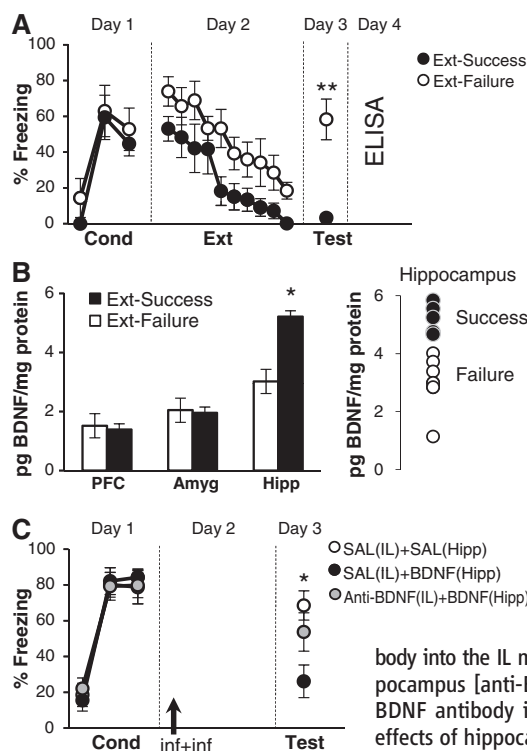


Fig. 3. Hippocampal projections to the infralimbic cortex mediate BDNF-extinction. (A) One day after extinction training, rats were divided into two groups on the basis of their ability to retrieve extinction (Ext-Success, $n = 6$; Ext-Failure, $n = 6$; $**P < 0.01$). The following day, BDNF protein concentrations were determined by enzyme-linked immunosorbent assay (ELISA). (B) Ext-Success rats showed elevated levels of BDNF in the hippocampus but not in the mPFC or amygdala ($*P < 0.05$, Ext-Success versus Ext-Failure). Hippocampal BDNF levels in individual Success and Failure rats were nonoverlapping. (C) Conditioned rats were divided into three groups. Controls received a saline infusion into the IL mPFC followed by a saline infusion into the hippocampus [SAL(IL) + SAL(Hipp), $n = 7$]. Another group received a saline infusion into the IL mPFC followed by a BDNF infusion into Hipp [SAL(IL) + BDNF(Hipp), $n = 8$]. A third group received an infusion of a BDNF-sequestering antibody into the IL mPFC followed by BDNF infusion into the hippocampus [anti-BDNF(IL) + BDNF(Hipp), $n = 9$]. Infusion of BDNF antibody into the IL mPFC blocked the fear-reducing effects of hippocampal BDNF. $*P < 0.05$, SAL(IL) + SAL(Hipp) compared to SAL(IL) + BDNF(Hipp).

amygdala (Fig. 3B). These data are consistent with previous studies in which genetic knockdown of hippocampal BDNF impaired fear extinction (17).

If the hippocampus is the source of IL BDNF, then increasing the available supply of hippocampal BDNF should have similar effects. We took advantage of the fact that BDNF infusions increase BDNF levels in efferent targets (18). There were three treatment groups in this experiment. After conditioning, one group received a hippocampal infusion of BDNF immediately after a saline infusion into the IL mPFC [SAL(IL) + BDNF(Hipp)]. A second group also received a hippocampal BDNF infusion, but this was preceded by infusion of a BDNF-inactivating antibody into the IL mPFC [anti-BDNF(IL) + BDNF(Hipp)] to test the hypothesis that Hipp-applied BDNF works via the IL mPFC. A control group received SAL infusions into both structures [SAL(IL) + SAL(Hipp)].

Similar to its effect on the IL mPFC, BDNF infused into the hippocampus reduced fear, as measured by both freezing [main effect of drug $F_{2,21} = 4.715$, $P = 0.020$, post hoc $P = 0.013$ comparing SAL(IL) + SAL(Hipp) to SAL(IL) + BDNF(Hipp)] (Fig. 3C) and conditioned suppression of food seeking (fig. S4). The effect of hippocampal BDNF could be prevented by coadministration of a BDNF-inactivating antibody in the IL mPFC [$P = 0.461$ comparing SAL(IL) + SAL(Hipp) to Anti-BDNF(IL) + BDNF(Hipp)], which suggests that the IL mPFC is the primary site of action for hippocampal BDNF.

We were able to pharmacologically induce extinction with a single infusion of BDNF into the hippocampal-infralimbic pathway, a key projection for extinction memory. This effect was not a facilitation of extinction, as no extinction training was required. We have adopted the term “BDNF-extinction” to parallel the term “BDNF-LTP” used to describe BDNF induction of hippocampal LTP in the absence of electrical stimulation (19). Extinction potentiates the hippocampal-prefrontal pathway, and disrupting this potentiation disrupts extinction recall (20). Our results provide further support for the importance of this pathway in extinction and extend these findings by identifying BDNF as a key molecular mediator.

In our experiments, BDNF-extinction required NMDA receptors, which are also necessary for extinction-related bursting in IL neurons (8). Because BDNF facilitates NMDA receptor currents (11, 12), exogenously applied BDNF may simulate extinction by inducing bursting in the IL mPFC. Additionally, BDNF-extinction may involve IL targets, such as intercalated (21) or basolateral amygdala (9, 15) neurons, which also participate in extinction.

Because the behavioral effects of BDNF were observed only when BDNF was infused after conditioning, it is possible that BDNF treatment may lead to partial reversal of conditioning-induced changes. Conditioning induces a rapid reduction in hippocampal BDNF, which reverts in 2 days (22). Extinction failure then may arise from a delayed normalization of BDNF levels after conditioning. If

so, application of BDNF to the hippocampus (or to the IL mPFC) may work to reduce fear by restoring BDNF to preconditioning levels and/or reversing conditioning-induced reductions in IL excitability (23).

Recall of extinction in healthy human subjects activates the ventromedial PFC and hippocampus (24), both of which are deficient in posttraumatic stress disorder (25). A single-nucleotide polymorphism in the gene encoding human BDNF (Val⁶⁶ → Met) results in extinction impairment (26) and decreases the release of BDNF from hippocampal neurons (27). Pharmacotherapies that increase hippocampal BDNF may prove to be efficacious treatments for fear disorders characterized by extinction impairments. BDNF-extinction is complementary to reconsolidation blockade, in which pharmacological agents are used to eliminate the original fear memory (7). Both approaches represent potentially powerful strategies to treat anxiety disorders by manipulating traumatic memories within fear circuits.

References and Notes

1. F. Sotres-Bayon, C. K. Cain, J. E. LeDoux, *Biol. Psychiatry* **60**, 329 (2006).
2. G. J. Quirk, D. Mueller, *Neuropsychopharmacology* **33**, 56 (2008).
3. M. R. Milad, G. J. Quirk, *Nature* **420**, 70 (2002).
4. F. Sotres-Bayon, L. Diaz-Mataix, D. E. Bush, J. E. LeDoux, *Cereb. Cortex* **19**, 474 (2009).
5. T. W. Bredy et al., *Learn. Mem.* **14**, 268 (2007).
6. C. R. Bramham, E. Messaoudi, *Prog. Neurobiol.* **76**, 99 (2005).
7. S. Duvarci, K. Nader, *J. Neurosci.* **24**, 9269 (2004).
8. A. Burgos-Robles, I. Vidal-Gonzalez, E. Santini, G. J. Quirk, *Neuron* **53**, 871 (2007).
9. W. A. Falls, M. J. Miserendino, M. Davis, *J. Neurosci.* **12**, 854 (1992).

10. E. Santini, R. U. Muller, G. J. Quirk, *J. Neurosci.* **21**, 9009 (2001).
11. I. B. Black, *J. Neurobiol.* **41**, 108 (1999).
12. E. S. Levine, R. A. Crozier, I. B. Black, M. R. Plummer, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10235 (1998).
13. C. Herry, R. Garcia, *J. Neurosci.* **22**, 577 (2002).
14. W. B. Hoover, R. P. Vertes, *Brain Struct. Funct.* **212**, 149 (2007).
15. J. P. Chhatwal, L. Stanek-Rattiner, M. Davis, K. J. Ressler, *Nat. Neurosci.* **9**, 870 (2006).
16. H. S. Phillips, J. M. Hains, G. R. Laramée, A. Rosenthal, J. W. Winslow, *Science* **250**, 290 (1990).
17. S. A. Heldt, L. Stanek, J. P. Chhatwal, K. J. Ressler, *Mol. Psychiatry* **12**, 656 (2007).
18. J. F. McGinty, T. W. Whitfield Jr., W. J. Berglund, *Brain Res.* **1314**, 183 (2010).
19. E. Messaoudi, S. W. Ying, T. Kanhema, S. D. Croll, C. R. Bramham, *J. Neurosci.* **22**, 7453 (2002).
20. R. Garcia, G. Spennato, L. Nilsson-Todd, J. L. Moreau, O. Deschaux, *Neurobiol. Learn. Mem.* **89**, 560 (2008).
21. E. Likhtik, D. Popa, J. Apergis-Schoute, G. A. Fidacaro, D. Paré, *Nature* **454**, 642 (2008).
22. A. M. Rasmussen, L. Shi, R. Duman, *Neuropsychopharmacology* **27**, 133 (2002).
23. E. Santini, G. J. Quirk, J. T. Porter, *J. Neurosci.* **28**, 4028 (2008).
24. R. Kalisch et al., *J. Neurosci.* **26**, 9503 (2006).
25. M. R. Milad et al., *Biol. Psychiatry* **66**, 1075 (2009).
26. F. Soliman et al., *Science* **327**, 863 (2010).
27. Z. Y. Chen et al., *Science* **314**, 140 (2006).
28. We thank V. Garcia, D. Merced, N. Padilla, C. Rodriguez, J. Rodriguez-Romaguera, D. Sierra-Mercado, and A. Torrado for technical assistance and D. Paré and J. F. McGinty for helpful comments. Supported by NIH grants MH058883, MH081975, NS043011, MH083516, MH085383, and RR003051 and by the University of Puerto Rico.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5983/1288/DC1
Materials and Methods

Figs. S1 to S5
References

11 January 2010; accepted 21 April 2010
10.1126/science.1186909

SphK1 Regulates Proinflammatory Responses Associated with Endotoxin and Polymicrobial Sepsis

Padmam Puneet,¹ Celestial T. Yap,¹ Lingkai Wong,² Lam Yulin,² Dow Rhoo Koh,¹ Shabbir Moochhala,³ Josef Pfeilschifter,⁴ Andrea Huwiler,⁵ Alirio J. Melendez^{1,6*}

During sepsis, activation of phagocytes leads to the overproduction of proinflammatory cytokines, causing systemic inflammation. Despite substantial information regarding the underlying molecular mechanisms that lead to sepsis, several elements in the pathway remain to be elucidated. We found that the enzyme sphingosine kinase 1 (SphK1) is up-regulated in stimulated human phagocytes and in peritoneal phagocytes of patients with severe sepsis. Blockade of SphK1 inhibited phagocyte production of endotoxin-induced proinflammatory cytokines. We observed protection against sepsis in mice treated with a specific SphK1 inhibitor that was enhanced by treatment with a broad-spectrum antibiotic. These results demonstrated a critical role for SphK1 in endotoxin signaling and sepsis-induced inflammatory responses and suggest that inhibition of SphK1 is a potential therapy for septic shock.

The incidence of sepsis, and death from septic shock, has increased over the past few decades (1, 2). During sepsis, the host's innate immune response to bacterial infection is primarily mediated by neutrophils and monocytes/macrophages (3). These cells express pattern-

recognition receptors (PRRs) that bind conserved molecular structures shared by groups of microorganisms (3). Upon stimulation, PRRs initiate inflammatory signaling pathways leading to secretion of proinflammatory mediators, which promote the elimination of infectious agents and the

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1290>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1290/DC1>

This article **cites 26 articles**, 11 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/328/5983/1290#otherarticles>

This article appears in the following **subject collections**:

Medicine, Diseases

<http://www.sciencemag.org/cgi/collection/medicine>

induction of tissue repair. Excessive production of inflammatory mediators, however, can induce septic shock by triggering a systemic inflammation, resulting in vascular leakage, tissue damage, multiple organ failure, and death (4, 5).

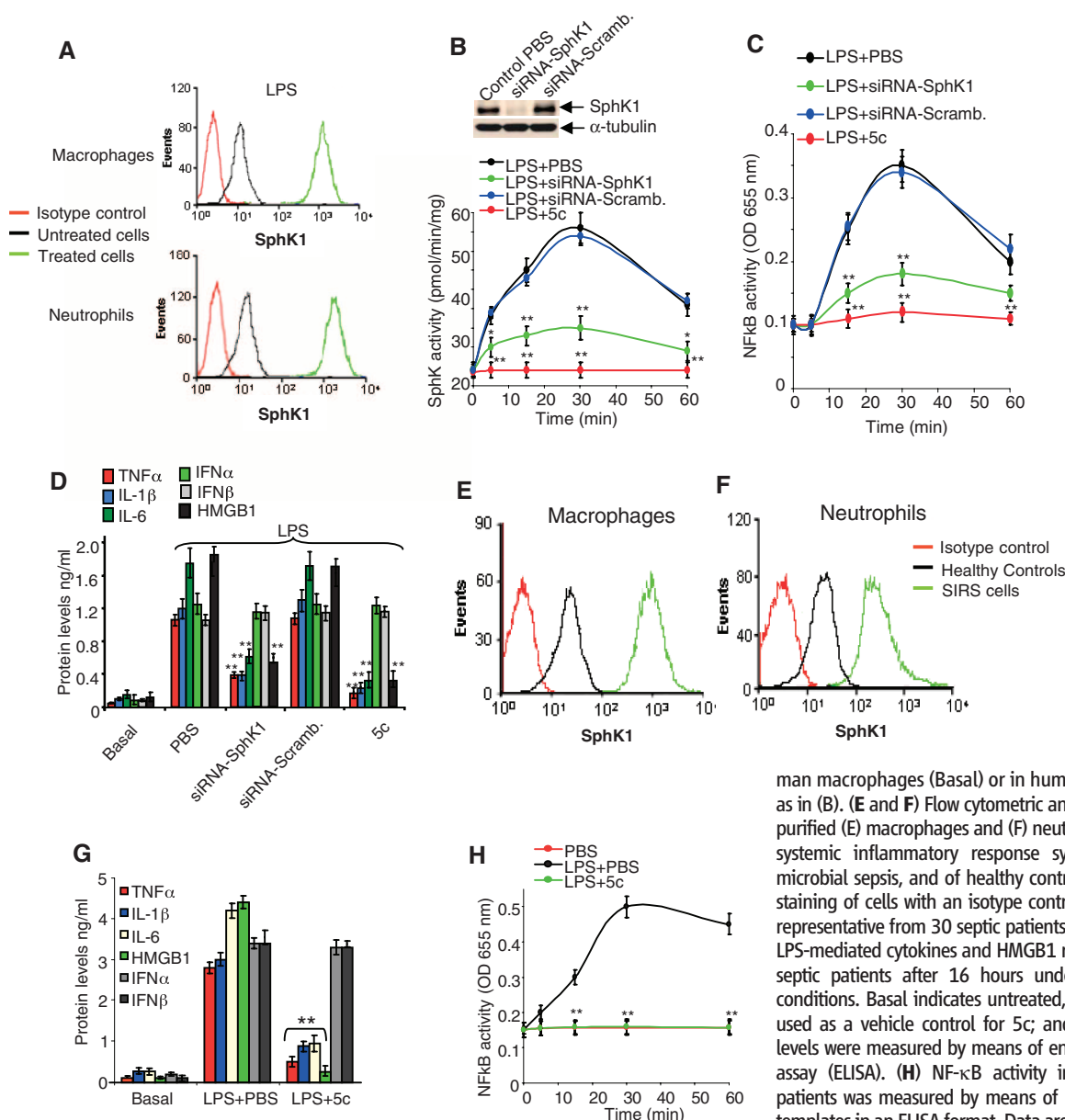
Several proinflammatory stimuli, including anaphylatoxin C5a, tumor necrosis factor- α (TNF- α),

and immune complexes, activate sphingosine kinase 1 (SphK1) on human neutrophils and macrophages, and blockade of SphK1 inhibits several proinflammatory responses triggered by these stimuli (6–12). SphKs are intracellular signaling enzymes that generate the lipid mediator sphingosine-1-phosphate (S1P) (13). The ability of SphK1 to mediate the secretion of proinflammatory mediators triggered by complement cascade anaphylatoxins (C5a) (9, 10) prompted us to investigate its role in inflammation caused by endotoxin and sepsis. We determined SphK1 protein and mRNA expression on human neutrophils and macrophages that were derived from peripheral blood and incubated in vitro with heat-inactivated Gram-positive bacteria, Gram-negative bacteria, mycobacteria, or with bacterial cell-wall components (14). SphK1 is constitutively expressed by most cells; however, SphK1 expres-

sion was strongly up-regulated by extracellular bacteria and lipopolysaccharide (LPS) (Fig. 1A) and bacterial lipoprotein (BLP) (fig. S1). In contrast, intracellular bacteria or the bacterial product, mycolic acid, had no effect (fig. S1). SphK1 was rapidly activated upon stimulation of human macrophages with LPS (Fig. 1B) or BLP (fig. S2). Silencing *SPHK1* by means of small interfering RNA (siRNA) (Fig. 1B), or inhibiting SphK1 with the specific SphK1 inhibitor 5c (fig. S3) (15), showed that SphK1 was required to activate the transcription factor nuclear factor κ B (NF- κ B) (Fig. 1C) and to induce the secretion of proinflammatory cytokines [TNF- α , interleukin-1 β (IL-1 β), and IL-6] and the proinflammatory protein high-mobility group protein B1 (HMGB1), but not type I interferon- α (IFN α) or IFN β (Fig. 1D).

¹Department of Physiology, National University of Singapore, 117597 Singapore. ²Department of Chemistry, National University of Singapore, 117543 Singapore. ³Defence Medical and Environmental Research Institute, DSO National Laboratories, 117510 Singapore. ⁴Pharmazentrum Frankfurt, University Hospital, Frankfurt am Main 60590, Germany. ⁵Institute of Pharmacology, University of Bern, Bern CH-3010, Switzerland. ⁶Medicine-Immunology, Infection and Inflammation, Glasgow Biomedical Research Centre, University of Glasgow, Glasgow G12 8TA, Scotland, UK.

*To whom correspondence should be addressed. E-mail: a.melendez-romero@clinmed.gla.ac.uk



separate experiments). Student's *t* test *P* values (**P* < 0.05 and ***P* < 0.01) are compared with LPS-induced control macrophages (LPS+PBS).

Fig. 1. Bacterial products increase human SphK1 expression and function in vitro. (A) Flow cytometric analysis of intracellular SphK1 expression after incubation of human macrophages and neutrophils with LPS and untreated control cells. Isotype control indicates staining of cells with an isotype control antibody. (B) SphK1 expression in human macrophages (top) pretreated with the siRNA-SphK1, scrambled siRNA (siRNA-Scramb.), or with vehicle control (Control PBS). SphK activity in human macrophages stimulated with LPS (bottom) in cells pretreated with vehicle control (PBS), siRNA-SphK1, siRNA-Scramb., or with the SphK1-inhibitor 5c. (C) NF- κ B activity in human macrophages stimulated as in (B). (D) Cytokine and HMGB1 production in untreated, nonstimulated human macrophages (Basal) or in human macrophages stimulated as in (B). (E and F) Flow cytometric analysis of SphK1 expression in purified (E) macrophages and (F) neutrophils from patients' aseptic systemic inflammatory response syndrome (SIRS) from polymicrobial sepsis, and of healthy controls. Isotype control indicates staining of cells with an isotype control antibody. Data shown are representative from 30 septic patients and 15 healthy samples. (G) LPS-mediated cytokines and HMGB1 release by macrophages from septic patients after 16 hours under the indicated stimulation conditions. Basal indicates untreated, unstimulated cells; PBS was used as a vehicle control for 5c; and the cytokines and HMGB1 levels were measured by means of enzyme-linked immunosorbent assay (ELISA). (H) NF- κ B activity in macrophages from septic patients was measured by means of p65 binding to specific DNA templates in an ELISA format. Data are shown as means $\pm$ SD (*n* = 4

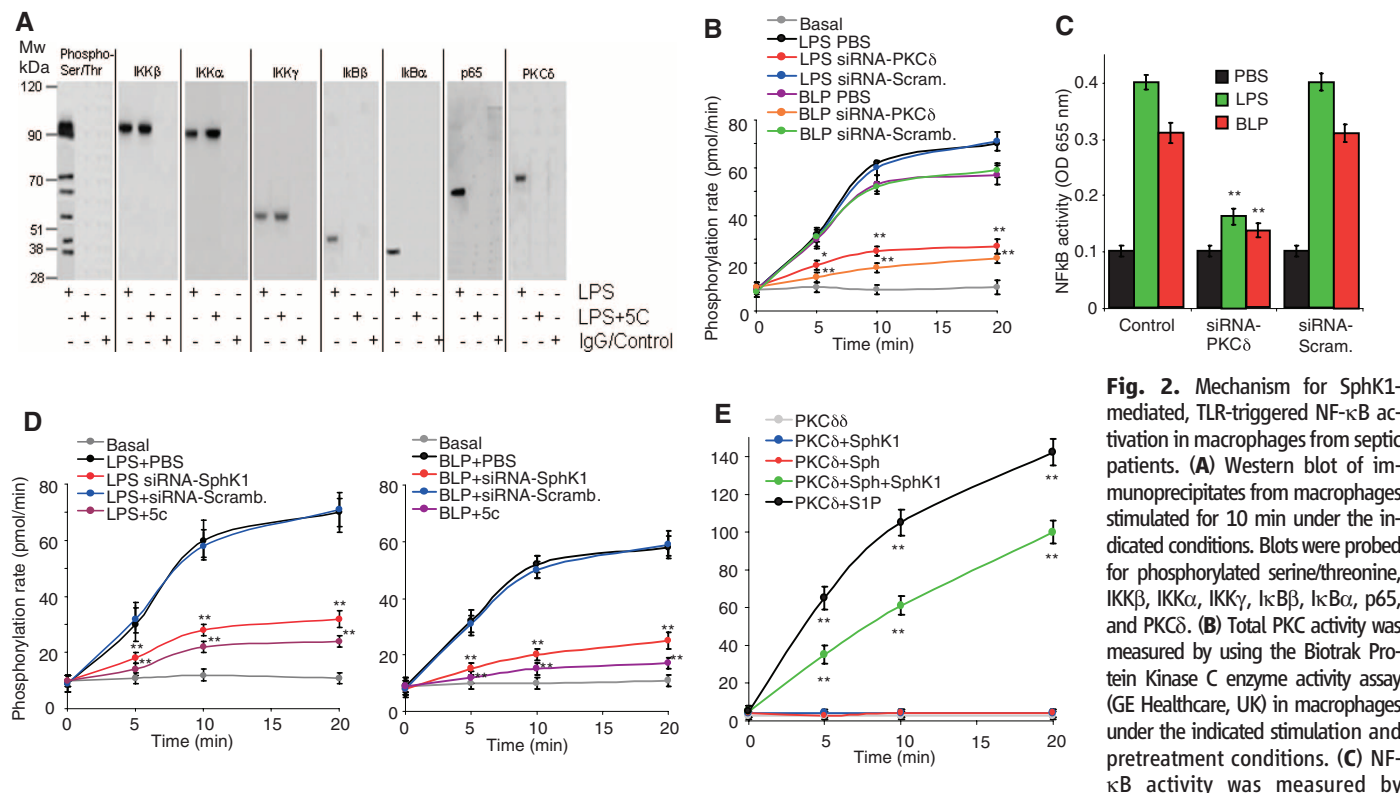
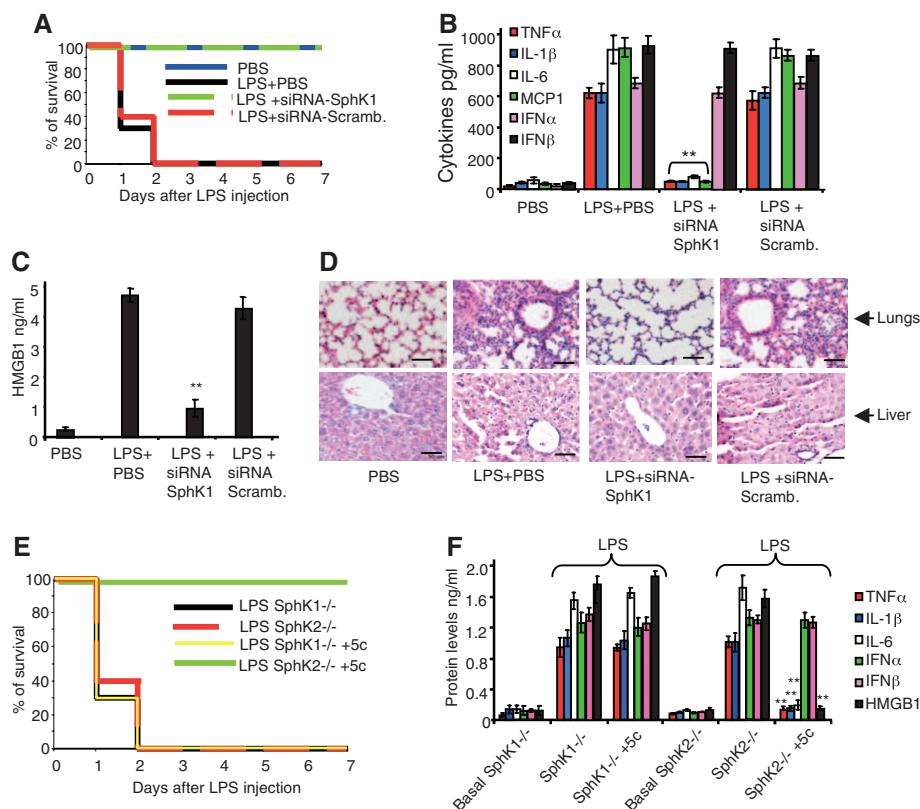


Fig. 2. Mechanism for SphK1-mediated, TLR-triggered NF- κ B activation in macrophages from septic patients. (A) Western blot of immunoprecipitates from macrophages stimulated for 10 min under the indicated conditions. Blots were probed for phosphorylated serine/threonine, IKK β , IKK α , IKK γ , I κ B β , I κ B α , p65, and PKC δ . (B) Total PKC activity was measured by using the Biotrak Protein Kinase C enzyme activity assay (GE Healthcare, UK) in macrophages under the indicated stimulation and pretreatment conditions. (C) NF- κ B activity was measured by

means of p65 binding to specific-DNA templates in an ELISA format in human macrophages under the indicated stimulation and pretreatment conditions. (D) Total PKC activity was measured using the Biotrak Protein Kinase C enzyme activity assay in macrophages stimulated with LPS or BLP under the indicated pretreatment conditions. Data are shown as means $\pm$ SD from multiple patients ($n = 10$ patient samples). Student's t test P values (** $P < 0.01$) are compared with LPS- or BLP-induced control macrophages (LPS+PBS or BLP+PBS). (E) PKC δ is directly activated by the SphK-catalyzed product S1P. Shown are PKC activity assay on recombinant PKC δ incubated with PBS (PKC δ +PBS), incubated with recombinant SphK1 (PKC δ +SphK1), incubated with Sphingosine (PKC δ +Sph), incubated with recombinant SphK1 and Sphingosine (PKC δ +Sph+SphK1), and incubated with S1P (PKC δ +S1P). Data are shown as means $\pm$ SD from triplicate measurements from three separate experiments. Student's t test P values (** $P < 0.01$) are compared with the control samples (PKC δ +PBS).

Fig. 3. Inhibition of SphK1 protects against LPS-induced endotoxic shock in mice. (A) Survival curves for LPS-induced death in mice pretreated with vehicle alone (LPS+PBS), SphK1-siRNA, or scrambled siRNA. As a negative control, mice were pretreated and challenged with vehicle alone (PBS). (B) Serum cytokines or (C) HMGB1 from LPS-injected mice pretreated under the indicated conditions, measured 24 hours after LPS challenge. (D) Immune-cell infiltration and tissue damage in the lungs and liver, 24 hours after LPS challenge, in mice pretreated under the indicated conditions. Immune cells were detected by means of haematoxylin and eosin staining; the magnification is $\times 10$. Scale bar, 30 μ m. (E) Survival curves for LPS-induced death in the indicated mouse strains in the presence or absence of pretreatment with 2 mg/kg of 5c. (F) LPS-triggered secretion of cytokines and HMGB1 from mouse macrophages from the indicated mouse strains in the presence or absence of pretreatment with 10 μ M 5c. Data points correspond to the mean $\pm$ SD of three independent experiments (six mice per treatment group; a total of 18 mice per condition were used). Student's t test P values (** $P < 0.01$) are compared with LPS-induced control (LPS+PBS).



We next studied the role of SphK1 in cells from patients with sepsis. We enrolled 30 patients with severe sepsis from different sources for this study. The main diagnoses included systemic *Staphylococcus aureus* infection in 12 (40%) patients, intra-abdominal infection in 12 (40%) patients, and trauma in six (20%) patients. Characteristics of the patients and the 15 healthy volunteer controls are summarized in table S1.

SphK1 expression was increased on infiltrating neutrophils and macrophages isolated from the peritoneal cavity of patients with septic shock (Fig. 1, E and F). Moreover, treating peritoneal macrophages with 5c blocked the LPS-triggered production of TNF- α , IL-1 β , IL-6, monocyte chemoattractant protein-1 (MCP-1), and HMGB1 but not type I interferons (Fig. 1G). LPS-triggered NF- κ B activation was also inhibited in patient-derived macrophages that had been pretreated with 5c (Fig. 1H). Furthermore, LPS triggered cytosolic production of S1P (fig. S4A), although we did not detect S1P in the supernatant of stimulated cells (fig. S4A). We obtained similar results using BLP (fig. S4B).

We next investigated which of the signaling molecules that were involved in the NF- κ B activation pathway downstream of Toll-like receptors (TLRs) were affected by SphK1 blockade. We

immunoprecipitated I κ B-kinase-beta (IKK β) from LPS-activated macrophages and probed for phosphoserine or threonine. Co-immunoprecipitation of several proteins peaked 10 min after LPS stimulation; however, pretreatment with 5c inhibited protein phosphorylation (Fig. 2A). We then probed for individual IKKs and members of the NF- κ B complex and found that most of the proteins pulled down by IKK β were part of IKK or NF- κ B complexes (IKK β , IKK α , IKK γ , I κ B β , I κ B α , or p65) (Fig. 2A). There was also a ~70 kD protein, which did not belong to the IKK or NF- κ B complexes. This protein was protein kinase C- δ (PKC δ) (Fig. 2A).

We next decided to establish whether PKC δ was indeed activated by LPS or BLP in macrophages and whether PKC δ played any role in the activation of NF- κ B. We silenced *PRKCD* expression by pretreating cells with a *PRKCD*-validated siRNA (PKC δ -siRNA) and analyzed the siRNA specificity on the expression of other PKC-isoforms (fig. S5A), including PKC ϵ , which is suggested to be involved in LPS-induced IKK and NF- κ B activation in monocytic cell lines (16). PKC δ -siRNA, but not control siRNA, largely inhibited LPS and BLP-triggered PKC activation (Fig. 2B). Furthermore, in cells pretreated with the PKC δ -siRNA, LPS- and BLP-triggered NF κ B activation was

substantially inhibited (Fig. 2C). siRNA-PKC δ -treated cells exhibited reduced phosphorylation of IKK α , IKK β , and IKK γ and impaired IKK β co-immunoprecipitation with I κ B α , I κ B β , and p65 (fig. S5B), providing further evidence that PKC δ may act downstream of SphK1 in human macrophages to trigger activation of NF- κ B.

We then investigated how SphK1 regulates PKC δ activation. We found that the LPS- and BLP-triggered PKC δ activation was inhibited in cells pretreated with the SphK1-siRNA or 5c (Fig. 2D). Moreover, in in vitro experiments in which we used recombinant PKC δ , we found that PKC δ can be activated by S1P (Fig. 2E). Taken together, these data suggest that LPS and BLP sequentially activate SphK1 and PKC δ , which then leads to the activation of IKK and NF- κ B and subsequent generation and release of proinflammatory molecules.

To assess the direct involvement of SphK1 in inflammatory responses, we tested whether in vivo silencing of *SPHK1* by means of siRNA or blocking SphK1 activity with 5c reduced systemic inflammation and lethal shock in mouse models of sepsis. A synthetic specific siRNA against SphK1 can silence SphK1 expression in vivo (17, 18), and we confirmed this in our system (fig. S5C). We then injected mice intraperitoneally with a lethal dose of

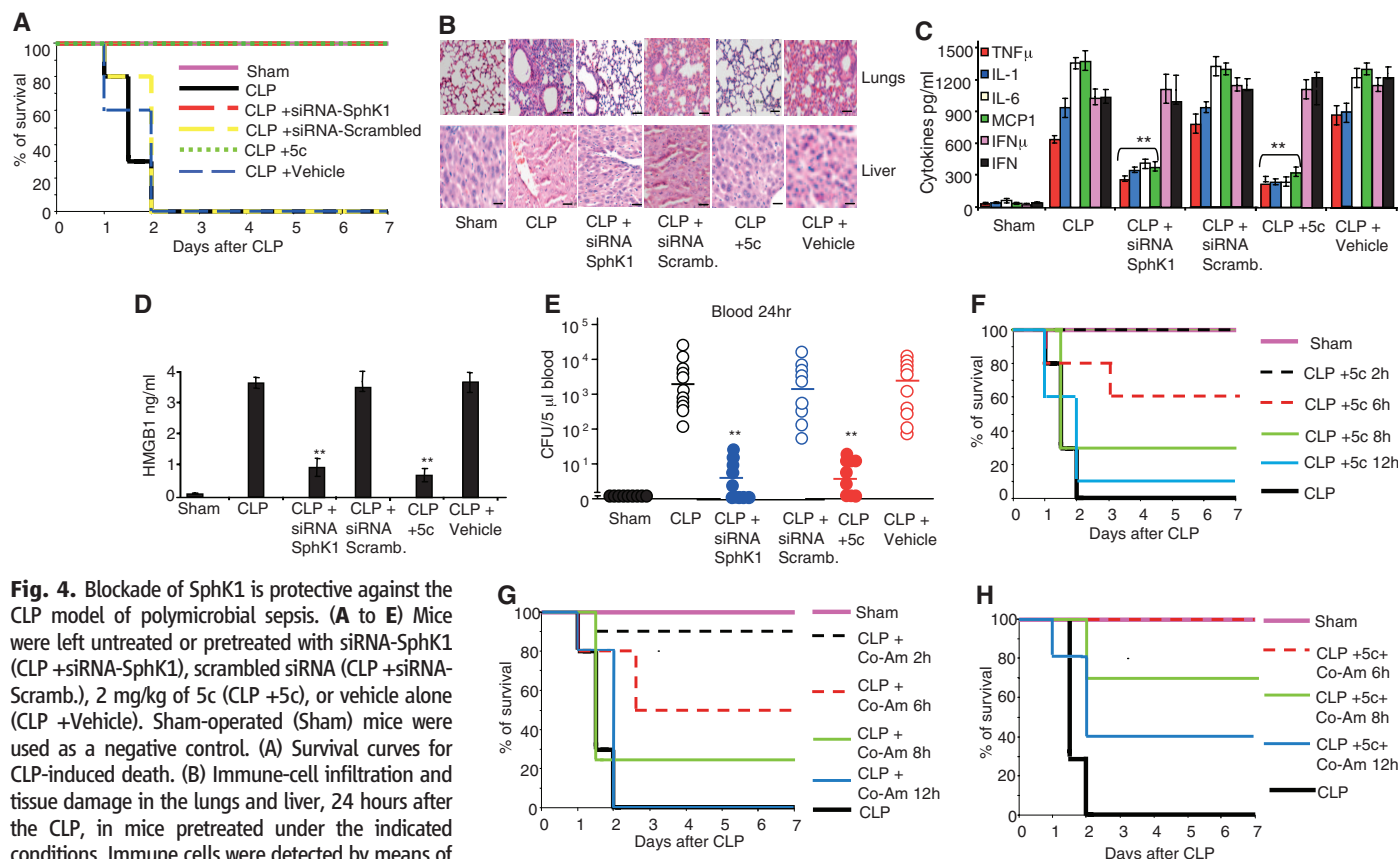


Fig. 4. Blockade of SphK1 is protective against the CLP model of polymicrobial sepsis. (A to E) Mice were left untreated or pretreated with siRNA-SphK1 (CLP + siRNA-SphK1), scrambled siRNA (CLP + siRNA-Scrambled), 2 mg/kg of 5c (CLP + 5c), or vehicle alone (CLP + Vehicle). Sham-operated (Sham) mice were used as a negative control. (A) Survival curves for CLP-induced death. (B) Immune-cell infiltration and tissue damage in the lungs and liver, 24 hours after the CLP, in mice pretreated under the indicated conditions. Immune cells were detected by means of haematoxylin and eosin staining; the magnification was $\times 10$. Scale bar, 30 μ m. (C) Serum cytokines were measured by means of ELISA 24 hours after CLP. (D) Serum HMGB1 measured 24 hours after CLP. (E) Bacteria in the blood were measured by the colony-forming unit method 24 hours after CLP. Student's *t* test *P* values (***P* < 0.01) are compared with CLP-induced control (CLP). (F) Survival

in mice treated with 5c (2 mg/kg) 2, 6, 8, or 12 hours after CLP. (G) Effects of the antibiotic co-amoxiclav (30 mg/kg) given after 2, 6, 8, or 12 hours on survival after CLP. (H) Survival of mice treated with 5c (2 mg/kg) and co-amoxiclav (30 mg/kg) 6, 8, and 12 hours after treatment. Data points correspond to the mean of three independent experiments (*n* = 6 mice per treatment group).

LPS in order to induce shock. Lethality was monitored over time and compared with mice that had been pretreated with saline alone (PBS), scrambled siRNA, or siRNA against SphK1 before LPS administration. We observed 100% survival of LPS-injected mice pretreated with siRNA-SphK1 (Fig. 3A).

Analysis of blood samples taken from mice pretreated with the siRNA-SphK1 after LPS administration showed a significant reduction of the plasma concentrations of TNF- α , IL-1 β , IL-6, MCP1 (Fig. 3B), and HMGB1 (Fig. 3C), whereas the amounts of IFN α and IFN β remained unchanged. We also observed a reduction in the total immune-cell infiltration into the liver and lungs of mice in which SphK1 had been silenced (Fig. 3D). To further study the protection provided by the blockade of SphK1, groups of mice were injected with increasing doses of 5c 30 min before the systemic injection of LPS. Survival was dose-dependent, ranging from no protection at 0.1 mg/kg to reaching 100% survival at 2mg/kg (fig. S6). Taken together, these results show that siRNA-silencing of *SPHK1* or SphK1 inhibition protects mice from endotoxic shock.

It has been reported that SphK1-null mice exhibit normal inflammatory cell recruitment during thioglycollate-induced peritonitis and in a collagen-induced arthritis (CIA) model of rheumatoid arthritis (19). In contrast and consistent with in vitro data (6–12), in vivo siRNA silencing of SphK1 in adult mice blocks C5a-induced acute peritonitis and chronic inflammation in the CIA model (17, 18), demonstrating that SphK1 is indeed required for acute and chronic inflammation. Thus, the phenotype of the mice in which SphK1 had been knocked out may be due to adaptive functional redundancy occurring during embryonic development (20, 21), necessitated by the fundamental roles of SphKs in cellular responses. Consistent with previous reports showing no protection from inflammatory stimuli (18), neither *SPHK1*^{−/−} nor *SPHK2*^{−/−} mice were protected from LPS-induced death and cellular responses (Fig. 3, E and F). When we treated both *SPHK1*^{−/−} and *SPHK2*^{−/−} mice with 5c, only the mice lacking SphK2 were protected (Fig. 3, E and F). These data demonstrate that SphK1 is indeed a critical player in endotoxic shock and confirm that 5c is a SphK1-specific inhibitor.

Experimental endotoxic shock reproduces human sepsis only in part because it does not involve the replication and dissemination of bacteria. Under these conditions, blockade of SphK1 signaling could be deleterious by impairing the capacity of the immune system to fight infections, as observed for treatments with antibody to TNF- α (22). We therefore investigated whether blockade of SphK1 protects against septic shock in a model of polymicrobial peritonitis and sepsis caused by cecal ligation and puncture (CLP), a model that resembles human microbial sepsis (23). We found that mice pretreated with the siRNA-SphK1 or with 5c were protected against CLP-induced systemic inflammation and mortality as compared with a control siRNA and vehicle control (Fig. 4A). Histological examination of the lungs and liver showed a substantial reduction

in inflammation and tissue damage (Fig. 4B) and lower amounts of inflammatory mediators (Fig. 4, C and D). Inhibition of SphK1 did not compromise bacterial clearance; in fact, blood bacterial load was lower in mice pretreated with the SphK1 inhibitor or with the siRNA-SphK1 (Fig. 4E). These results suggest that blockade of SphK1 reduces inflammatory responses and may aid in control of bacterial infection. Use of *SPHK1*^{−/−} and *SPHK2*^{−/−} mice for the CLP model generated similar results as those for the endotoxin-mediated shock shown in Fig. 3, E and F (fig. S7).

To assess the therapeutic potential of SphK1 blockade in sepsis, we monitored whether blockade of SphK1 was still protective when 5c was administered 2, 6, 8, and 12 hours after the CLP procedure. 5c conferred 100% protection when applied 2 hours after CLP, 60% when applied 6 hours after CLP, 30% when applied 8 hours after CLP, and 10% when applied 12 hours after CLP (Fig. 4F). Thus, 5c is effective as a therapeutic for experimental sepsis.

We next treated CLP septic mice with 5c and a broad-spectrum antibiotic, co-amoxiclav, which is currently used to treat patients with sepsis. Administration of a broad-spectrum antibiotic is protective when administered before or shortly after the CLP procedure; however, the time-window for treatment opportunity is very narrow, and in the clinic, despite antibiotic treatment, a large number of patients with sepsis or septic shock still die. Co-amoxiclav alone increased survival by 50% when administered 6 hours after CLP and by 25% when administered 8 hours after CLP, and no survival was observed when it was administered 12 hours after CLP (Fig. 4G). Co-amoxiclav-mediated survival correlated with blood-bacterial clearance (fig. S8). The combination of 5c with co-amoxiclav enhanced survival to 100% when co-injected 6 hours after CLP, to 70% when co-injected 8 hours after CLP, and to 40% when the combination therapy was administered 12 hours after CLP (Fig. 4H). Thus, our data demonstrate a clear advantageous synergy with this combination therapy.

Studies in models of endotoxin-induced acute lung injury suggest that SIP and its receptors are critical players in maintaining the integrity of the vascular barrier function (24, 25). Our results suggest that, in the endotoxin and CLP shock models, elevated cytosolic SphK1 activity leads to an aberrant inflammatory response, resulting in multiple organ damage and death. Our results also suggest that blockade of SphK1 during septic shock does not interfere with the systemic SIP gradient that is required to maintain the integrity of the vascular barrier function.

SphK1 inhibition shows a clear potential therapeutic advantage against sepsis as compared with treatments with known anti-inflammatory agents, including TNF- α monoclonal antibodies and nitric oxide synthase inhibitors, which increase lethality potentially through suppression of the host's ability to fight infection (22, 26). Indeed, bacterial clearance is enhanced when SphK1 is inhibited. Research efforts have recently been directed at finding a mech-

anism for limiting the excessive inflammation triggered by TLRs as a way of providing potential novel therapies to treat sepsis and other inflammatory disorders (27). Here, we provide such a mechanism by targeting SphK1. Thus, administration of SphK1 inhibitors—alone or in combination with antibiotics, after the initiation of sepsis—might offer a suitable new therapeutic tool for the treatment of septic shock and other microbe-mediated diseases in humans, in whom “out-of-control” inflammation often leads to fatal outcomes.

References and Notes

- G. S. Martin, D. M. Mannino, S. Eaton, M. Moss, *N. Engl. J. Med.* **348**, 1546 (2003).
- R. Phillips Dellinger, J. E. Parrillo, *Crit. Care Med.* **32**, 282 (2004).
- C. A. Janeway Jr., R. Medzhitov, *Annu. Rev. Immunol.* **20**, 197 (2002).
- A. L. Beal, F. B. Cerra, *JAMA* **271**, 226 (1994).
- P. E. Marik, J. Varon, *Dis. Mon.* **47**, 465 (2001).
- A. Melendez et al., *Curr. Biol.* **8**, 210 (1998).
- A. J. Melendez, L. Bruetsch, R. A. Floto, M. M. Harnett, J. M. Allen, *Blood* **98**, 3421 (2001).
- A. J. Melendez, A. K. Khaw, *J. Biol. Chem.* **277**, 17255 (2002).
- F. B. Ibrahim, S. J. Pang, A. J. Melendez, *J. Biol. Chem.* **279**, 44802 (2004).
- A. J. Melendez, F. B. Ibrahim, *J. Immunol.* **173**, 1596 (2004).
- L. Zhi, B. P. Leung, A. J. Melendez, *J. Cell. Physiol.* **208**, 109 (2006).
- A. J. Melendez et al., *Nat. Med.* **13**, 1375 (2007).
- A. J. Melendez, *Biochim. Biophys. Acta* **1784**, 66 (2008).
- Materials and methods are available as supporting material on Science Online.
- L. Wong, S. S. Tan, Y. Lam, A. J. Melendez, *J. Med. Chem.* **52**, 3618 (2009).
- A. Castrillo et al., *J. Exp. Med.* **194**, 1231 (2001).
- P. N. Pushparaj, S. C. H'ng, A. J. Melendez, *Int. J. Biochem. Cell Biol.* **40**, 1167 (2008).
- J. Michaud, M. Kohn, R. L. Proia, T. Hla, *FEBS Lett.* **580**, 4607 (2006).
- W. Q. Lai et al., *J. Immunol.* **181**, 8010 (2008).
- X. Gu, *Trends Genet.* **19**, 354 (2003).
- R. A. Kafri, A. Bar-Even, Y. Pilpel, *Nat. Genet.* **37**, 295 (2005).
- M. K. Eskandari et al., *J. Immunol.* **148**, 2724 (1992).
- W. J. Hubbard et al., *Shock* **24**, (Suppl 1), 52 (2005).
- B. J. McVerry, J. G. Garcia, *Cell. Signal.* **17**, 131 (2005).
- R. Wadgaonkar et al., *Am. J. Physiol. Lung Cell. Mol. Physiol.* **296**, L603 (2009).
- P. Q. Eichacker et al., *Am. J. Respir. Crit. Care Med.* **166**, 1197 (2002).
- H. Weighardt, B. Holzmann, *Immunobiology* **212**, 715 (2008).
- We thank A.-K. Melendez-Fraser for proofreading the manuscript. We thank M. M. Harnett for advice during the preparation of the manuscript. This work was supported by grants from the Medical Research Council UK, the Biomedical Research Council of Singapore, the Swiss National Foundation (to A.H.) and the German Research Foundation (to A.H. and J.P.). A patent application has been filed for compound 5c (U.S. Provisional Patent Application 61/217,334) by A.J.M., Y.L., X.H., and L.W. A material transfer agreement is required for use of *SphK1*^{−/−} and *SphK2*^{−/−} mice and compound 5c.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5983/1290/DC1
Materials and Methods
Figs. S1 to S8
Table S1
References

22 February 2010; accepted 19 April 2010
10.1126/science.1188635

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1295>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/328/5983/1295/DC1>

This article **cites 16 articles**, 7 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/328/5983/1295#otherarticles>

This article appears in the following **subject collections**:

Microbiology

<http://www.sciencemag.org/cgi/collection/microbio>

Techniques

<http://www.sciencemag.org/cgi/collection/techniques>

Asymmetrical Distribution of the Second Messenger c-di-GMP upon Bacterial Cell Division

Matthias Christen,¹ Hemantha D. Kulasekara,¹ Beat Christen,² Bridget R. Kulasekara,³ Lucas R. Hoffman,⁴ Samuel I. Miller^{1,5*}

The bacterial second messenger cyclic diguanosine monophosphate (c-di-GMP) regulates cellular motility and the synthesis of organelles and molecules that promote adhesion to a variety of biological and nonbiological surfaces. These properties likely require tight spatial and temporal regulation of c-di-GMP concentration. We have developed genetically encoded fluorescence resonance energy transfer (FRET)-based biosensors to monitor c-di-GMP concentrations within single bacterial cells by microscopy. Fluctuations of c-di-GMP were visualized in diverse Gram-negative bacterial species and observed to be cell cycle dependent. Asymmetrical distribution of c-di-GMP in the progeny correlated with the time of cell division and polarization for *Caulobacter crescentus* and *Pseudomonas aeruginosa*. Thus, asymmetrical distribution of c-di-GMP was observed as part of cell division, which may indicate an important regulatory step in extracellular organelle biosynthesis or function.

Cyclic diguanosine monophosphate (c-di-GMP) has been implicated in regulating a variety of bacterial cellular characteristics, including antibiotic resistance, biofilm formation, extracellular carbohydrate and adhesin production, pilus- and flagellum-based motility, and cell

cycle progression (1–3). Multiple diguanylate cyclases (DGCs) and phosphodiesterases (PDEs) synthesize and degrade c-di-GMP. Regulation of biological functions is accomplished by c-di-GMP's binding to a diverse array of receptors, including PilZ domain proteins (4, 5), cyclic diguanylate

riboswitches (6), and transcription factors (7, 8). The apparent role of c-di-GMP in the cell cycle and the presence of many paralogous DGC, PDE, and PilZ proteins controlling diverse cellular functions indicate that there is likely tight spatial and temporal regulation of c-di-GMP.

To study the spatiotemporal dynamics of c-di-GMP fluctuations in individual living bacterial cells, we engineered a set of genetically encoded fluorescence resonance energy transfer (FRET)-based biosensors by fusing PilZ proteins between cyan and yellow fluorescent proteins (CFP and YFP, respectively) (fig. S1) (9). The FRET response of each biosensor to various c-di-GMP concentrations was characterized in vitro (table S1). Binding of c-di-GMP to the diguanylate recep-

¹Department of Immunology, University of Washington, Seattle, WA 98195, USA. ²Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305, USA. ³Molecular and Cellular Biology Program, University of Washington, Seattle, WA 98195, USA. ⁴Department of Pediatrics, University of Washington, Seattle, WA 98195, USA. ⁵Departments of Genome Sciences, Microbiology, and Medicine, University of Washington, Seattle, WA 98195, USA.

*To whom correspondence should be addressed. E-mail: millersi@uw.edu

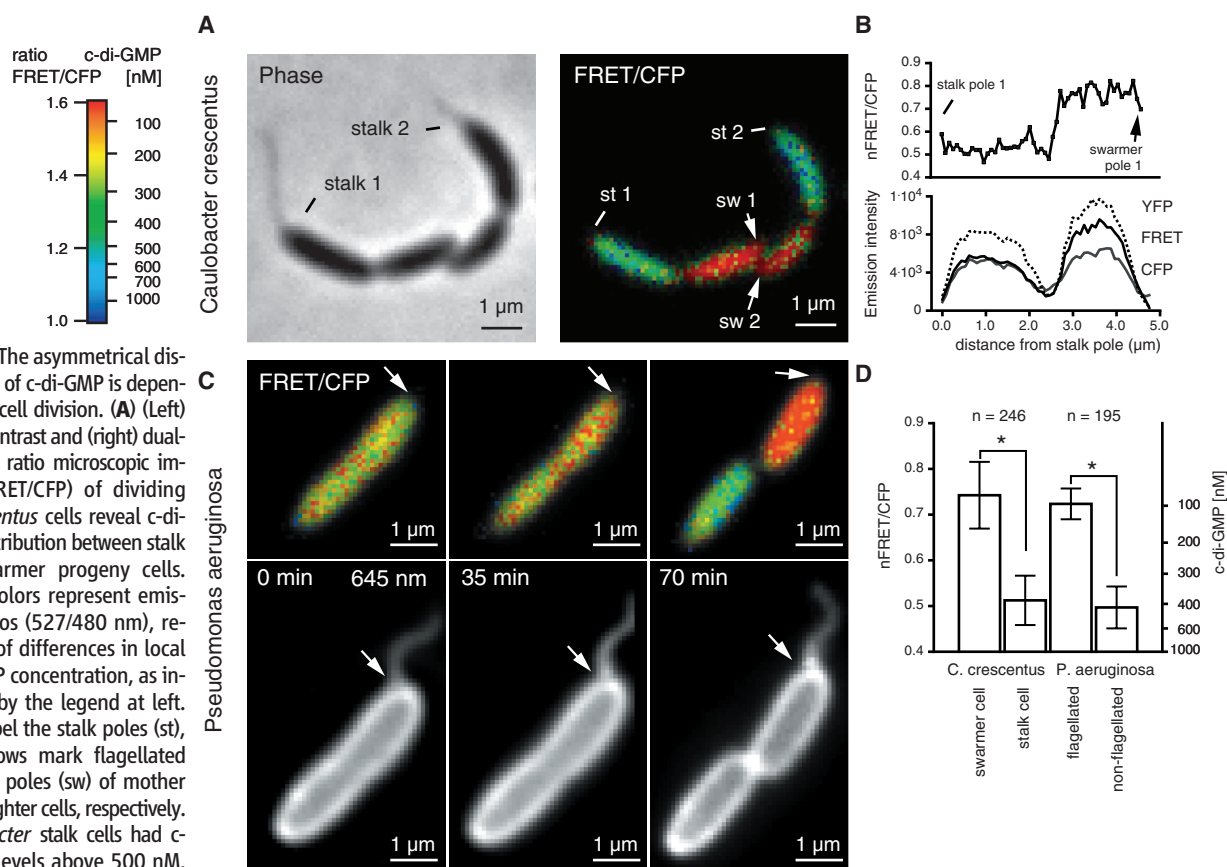


Fig. 1. The asymmetrical distribution of c-di-GMP is dependent on cell division. (A) (Left) Phase-contrast and (right) dual-emission ratio microscopic images (FRET/CFP) of dividing *C. crescentus* cells reveal c-di-GMP distribution between stalk and swarmer progeny cells. Pseudocolors represent emission ratios (527/480 nm), reflective of differences in local c-di-GMP concentration, as indicated by the legend at left. Lines label the stalk poles (st), and arrows mark flagellated swarmer poles (sw) of mother and daughter cells, respectively. *Caulobacter* stalk cells had c-di-GMP levels above 500 nM, whereas levels were below 100 nM in the swarmer cell. (B) (Top) Profile of the ratio of net FRET (FRET – spectral bleed-through – direct YFP excitation) to CFP emission fluorescence and (bottom) emission intensity, along the stalk-swarmer polar cell axis of the dividing *Caulobacter* cell no. 1 from (A). (C) C-di-GMP polarization in a dividing *P. aeruginosa* cell. (Top) Dual-emission ratio microscopic images

(FRET/CFP). (Bottom) Cells have been stained with Alexa Fluor 594; arrows indicate the position of the polar flagellum. (D) Mean nFRET/CFP ratios for flagellated and nonflagellated *Caulobacter* and *Pseudomonas* cells as indicated. Results shown are averages of single-cell measurements from at least three independent fields $\pm$ SD, * $P < 0.005$.

Fig. 2. C-di-GMP signaling patterns in wild-type and mutant *Caulobacter* strains. **(A)** Dual-emission ratio FRET images of *Caulobacter* cells of a wild-type strain, a strain expressing a constitutively active diguanylate cyclase (*pVan::dgcA*), or *pleC* and *pleD* deletion mutants in log-arithmically growing cultures. Pseudocolors represent emission ratios (FRET/CFP, 527/480 nm), which correlate to cellular c-di-GMP levels. Lines indicate stalk poles (st), and arrows indicate swarmer poles (sw). **(B)** Dot plots illustrating distributions of cellular c-di-GMP levels in logarithmically growing cultures of the indicated *Caulobacter* strains. Each data point in the graphs represents an individual bacterial cell for which the mean donor channel (CFP) signal is plotted against the mean net FRET (nFRET) intensity. Cells with relatively low c-di-GMP (<100 nM) are in the upper left half of each plot; cells with elevated c-di-GMP levels (>500 nM) are in the lower right. **(C)** Histogram showing the percentage of cells of the indicated strains from logarithmic growing cultures with a c-di-GMP concentration below 100 nM. For each condition, the average of three independent measurements $\pm$ SD is plotted. **(D)** Histogram depicting the nFRET/CFP ratio distribution in the swarmer cell and stalk cell subpopulation during the S to G₂ transition of cells of the indicated *Caulobacter* strains.

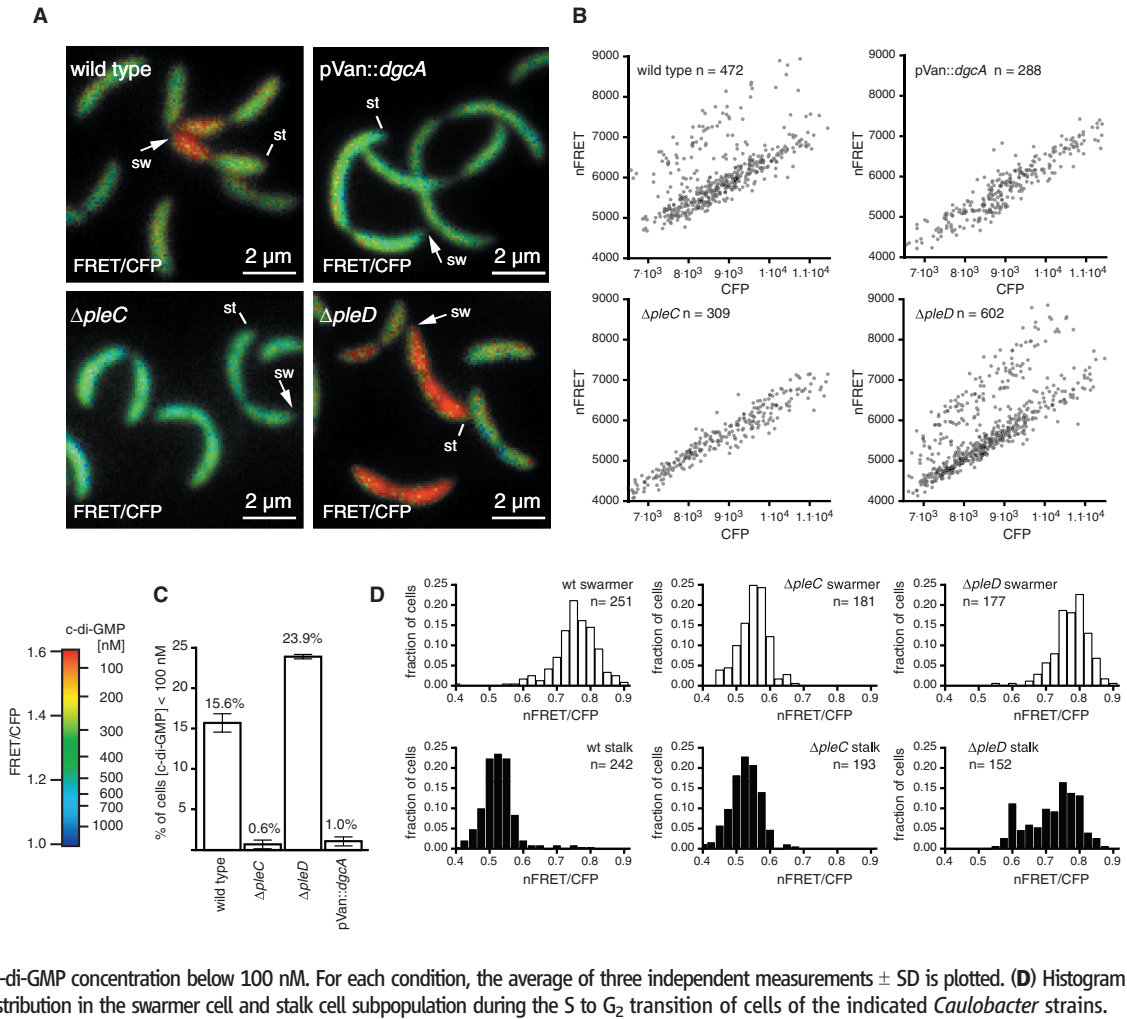
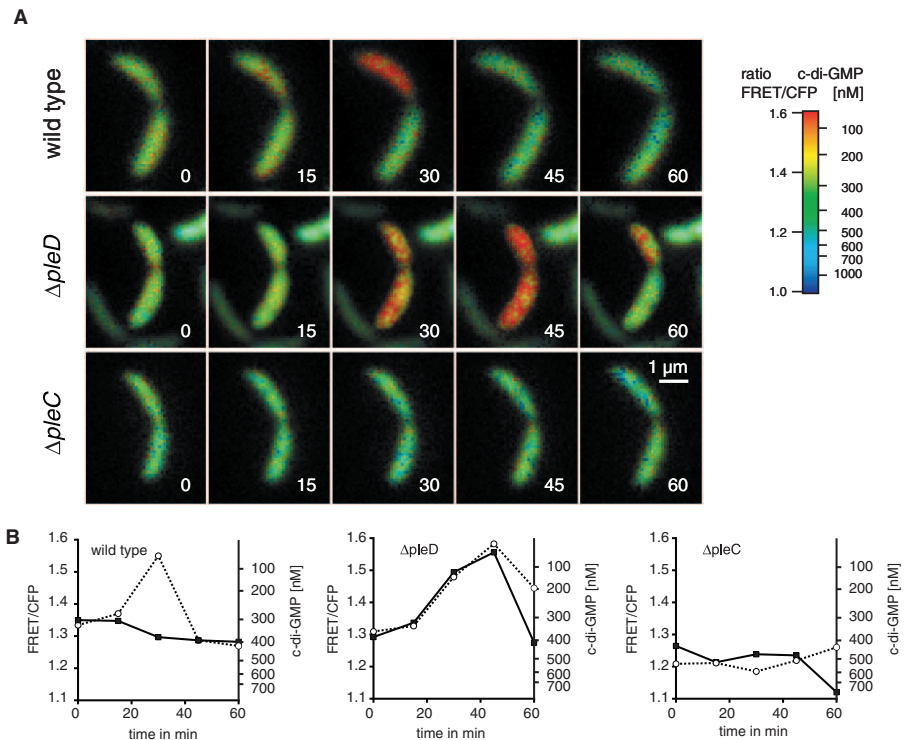


Fig. 3. Kinetics of fluorescence ratio changes (527/480 nm), reflecting c-di-GMP levels recorded in *Caulobacter* cells during the S to G₂ transition. **(A)** Time-lapse dual-emission ratiometric FRET microscopy of representative cells of the indicated strains recorded at intervals of 15 min (see also movies S2 to S4). **(B)** Corresponding plots of the emission ratio changes for the indicated strains over time in the swarmer cell (open circles) and the stalk cell (filled squares).



tor domain induced a conformational change (4, 10) that altered the relative orientation of the external fluorescent subunits, which decreased FRET efficiency. Thus, the fluorescence properties of these biosensors (i.e., the FRET/CFP emission ratio) could reflect cellular c-di-GMP levels. The biosensor derived from *Salmonella enterica* serovar Typhimurium protein YcgR exhibited the largest change in net FRET (nFRET) in vitro (−60.6%), with a binding constant of 195 nM for c-di-GMP and no detectable response to cyclic adenosine 3',5'-monophosphate, cyclic GMP, or guanosine 5'-triphosphate (figs. S1 to S3).

For live-cell imaging of c-di-GMP signaling events in single bacterial cells, we expressed the YcgR-based c-di-GMP sensor in *Caulobacter crescentus* and measured FRET efficiencies by ratiometric dual-emission microscopy (Fig. 1, A and B). *Caulobacter* has asymmetric cell division, resulting in two morphologically distinct progeny: a surface-attached replication-competent stalk cell, and a motile swarmer cell with a polar flagellum. The two daughter cells exhibited different emission ratios, with an average nFRET/CFP ratio of 0.7424 ± 0.073 for swarmer cells and 0.5124 ± 0.054 for stalk cells. Correlating the emission ratio (FRET/CFP) with the binding curve derived from in vitro characterization (fig. S2) yielded a cytosolic c-di-GMP concentration below 100 nM for the flagellated swarmer cell and a concentration above 500 nM for the stalk cell (Fig. 1, A and B, and Fig. 2D). Thus, c-di-GMP is asymmetrically distributed in *Caulobacter* immediately after cell division, with low apparent levels in the flagellated swarmer cell and levels at least five times as high in the nonmotile stalk cell.

Like *Caulobacter*, *Pseudomonas* produces a single polar flagellum that is partitioned to one daughter at cell division and can be visualized microscopically by labeling living cells with Alexa Fluor 594 conjugated to surface amine-specific succinimidyl ester (Fig. 1C, figs. S5 and S6, and movie S1). Before cell division, the two cell poles harbored similar c-di-GMP levels, probably because of rapid diffusion of the second-messenger molecule. When septum formation resulted in two distinct cells, c-di-GMP rose above 500 nM in the nonflagellated daughter cell and dropped in the flagellated cell below 100 nM (Fig. 1C). This asymmetrical second-messenger distribution was not a stochastic event: In both organisms, c-di-GMP levels were always significantly lower in the flagellated cell than in the nonflagellated one (Fig. 1D).

Bacterial genomes encode multiple DGCs and PDEs in proteins with signal-sensing domains (fig. S4), some of which exhibit distinct subcellular localization (11–13). Thus, asymmetrical distribution of c-di-GMP might be caused by spatially restricted expression or activation of individual DGC and PDE enzymes within each daughter cell. We reasoned that overexpression of a DGC would abolish this asymmetrical distribution by increasing c-di-GMP, and we there-

fore compared the cellular c-di-GMP distribution patterns in various *Caulobacter* strains. In a strain expressing the constitutively active DGC DgcA (14), c-di-GMP concentrations in the swarmer cells also rose (Fig. 2), which indicated that a localized decrease in DGC activity could cause the drop in c-di-GMP in wild-type swarmer cells. One likely candidate for a DGC that becomes inactivated in the swarmer cell is the diguanylate cyclase PleD. The DGC output activity of PleD is regulated via phosphorylation of its N-terminal receiver domain (11) by the histidine kinase DivJ and the phosphatase PleC, which localize to opposite poles of the predivisional cell (15). Upon cell division, PleC becomes restricted to the swarmer cell, where it dephosphorylates and subsequently inactivates PleD (3). Consistent with this model, a $\Delta pleC$ mutant yielded equivalent nFRET/CFP ratios for the swarmer cell (0.5318 ± 0.042) and for the stalked cell (0.5317 ± 0.046) (Fig. 2, A and D), which indicated that *pleC* is normally required to lower c-di-GMP concentrations in swarmer cells. In contrast, for a $\Delta pleD$ mutant, the average nFRET/CFP ratio in the stalk cell increased to 0.7057 ± 0.076 , whereas the swarmer cell maintained its normally high ratio of 0.7631 ± 0.051 (Fig. 2, A and D), which resulted in c-di-GMP concentrations below 100 nM in both cells.

Time-lapse experiments using wild-type and mutant *Caulobacter* strains further defined c-di-GMP dynamic changes during cell division (for parallel measurements in *Pseudomonas* see figs. S5 and S6 and movie S1). Wild-type *Caulobacter* cells before cell division had equivalent c-di-GMP levels (Fig. 3A and movie S2), and upon septation, c-di-GMP decreased in the flagellated swarmer cell below 100 nM (Fig. 3, A and B). Note that the swarmer cell-specific drop in c-di-GMP was a transient phenomenon, and the levels of c-di-GMP reverted within less than 20 min to those found in the stalk cell. In contrast, the $\Delta pleD$ mutant demonstrated a drop in both cells during the transition from S to G₂ phases and extended the period of decreased c-di-GMP levels in the swarmer cell (Fig. 3, A and B, and movie S3). Consistent with PleC function as a negative regulator of PleD activity, cellular c-di-GMP levels remained constitutively high in $\Delta pleC$ cells throughout the cell cycle (Fig. 3, A and B, and movie S4). Thus, *pleC* is required for asymmetrical distribution of c-di-GMP during cell division, most likely by spatially restricting PleD phosphorylation, and its DGC activity, to the stalk cell. Nevertheless, deletion of *pleD* itself did not result in a complete loss of cellular c-di-GMP, which suggested that additional DGC enzymes contribute to cell cycle-dependent c-di-GMP fluctuations.

Impairing the cellular distribution of c-di-GMP has major consequences for *Caulobacter* polar organelle development and function. Deletion of *pleC* and overexpression of *dgcA* both lead to a nonmotile phenotype, most likely as a result of a generalized increase in c-di-GMP and

activation of DgrA, which can in turn inhibit motility by binding to flagellar cytoplasmic components (4, 5, 16). Conversely, a *pleD* deletion mutant is hypermotile and expresses functional flagella at both cell poles (2), which indicates that the observed drop in c-di-GMP at cell division (Fig. 3) promotes rapid motility of flagellated cells immediately after septum formation. In addition to their use in *Pseudomonas* and *Caulobacter*, we have successfully used FRET-based c-di-GMP biosensor constructs in peritrichously flagellated *Salmonella enterica* serovar Typhimurium and nonflagellated *Klebsiella pneumoniae*. We observed that these organisms also exhibit asymmetric second-messenger distribution upon cell division (fig. S7), which suggests that this phenomenon is not unique to *Pseudomonas* and *Caulobacter*, and cell properties other than motility are regulated by asymmetrical second-messenger distribution upon cell division.

References and Notes

1. L. R. Hoffman et al., *Nature* **436**, 1171 (2005).
2. G. B. Hecht, A. Newton, *J. Bacteriol.* **177**, 6223 (1995).
3. P. Aldridge, R. Paul, P. Goymer, P. B. Rainey, U. Jenal, *Mol. Microbiol.* **47**, 1695 (2003).
4. M. Christen et al., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4112 (2007).
5. K. Paul, V. Nieto, W. C. Carlquist, D. F. Blair, R. M. Harshey, *Mol. Cell* **38**, 128 (2010).
6. N. Sudarsan et al., *Science* **321**, 411 (2008).
7. J. L. Leduc, G. P. Roberts, *J. Bacteriol.* **191**, 7121 (2009).
8. P. V. Krasteva et al., *Science* **327**, 866 (2010).
9. T. Ohashi, S. D. Galiacy, G. Briscoe, H. P. Erickson, *Protein Sci.* **16**, 1429 (2007).
10. J. Benach et al., *EMBO J.* **26**, 5153 (2007).
11. C. Chan et al., *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17084 (2004).
12. Z. T. Güvener, C. S. Harwood, *Mol. Microbiol.* **66**, 1459 (2007).
13. E. Huitema, S. Pritchard, D. Matteson, S. K. Radhakrishnan, P. H. Viollier, *Cell* **124**, 1025 (2006).
14. B. Christen et al., *J. Biol. Chem.* **281**, 32015 (2006).
15. R. T. Wheeler, L. Shapiro, *Mol. Cell* **4**, 683 (1999).
16. A. Boehm et al., *Cell* **141**, 107 (2010).
17. We thank R. Jones from Rutgers University, Piscataway, NJ, for c-di-GMP and T. Ohashi from Duke University, Durham, NC, for the pET15b::mCYPet_12AA_mYPet plasmids. This research was supported by the National Institute of Allergy and Infectious Diseases, NIH (1R21NS067579-01 and U54AI057141 to S.I.M.; KO8AI066251 to L.R.H. and 1R21NS067579-01 to H.D.K.), NSF (Graduate Research Fellowship to B.R.K.), Swiss National Foundation (PA00P3-124140 and PBBSA-120489 to M.C. and PA00P3-126243 to B.C.), Novartis Foundation to M.C., and Cystic Fibrosis Foundation (HOFFMA04L0 to L.R.H.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5983/1295/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 to S3
References
Movies S1 to S4

23 February 2010; accepted 23 April 2010
10.1126/science.1188658

NEW PRODUCTS



CELL IMAGER

The Cell Imager is for automated cell growth monitoring in the CellHost culture system. The Cell Imager accurately measures cell confluence and growth in 96-well plates and is suitable for measuring hybridoma clones in antibody screening for therapeutic and diagnostic research and development. It optimizes workflow by providing timely information with which to make decisions on media changes and culture steps. It can be placed directly on the Microlab Star workstation deck, and imaging and pipetting are done in the same position for maximum space economy. It reads a 96-well plate in about five minutes, with its two positions allowing one plate to be imaged while another is processed. Its software provides live data tracking for a simultaneous overview of cell confluence in all plate wells.

Hamilton Robotics

For info: +49-89-552-64-915 | www.hamiltonrobotics.com

LC/MS SYSTEM

The combination of the Exactive liquid chromatography/mass spectrometry (LC/MS) system with Transcend technology provides a complete LC/MS solution. This configuration combines automated online sample preparation for high throughput with the dual benefits of ultrahigh resolution and accurate mass analysis for challenging screening applications, including complex samples containing pesticides or metabolites. Leveraging the proven mass analyzer technology of the Orbitrap platform, the Exactive delivers ultrahigh resolution and accurate mass data to enable the rigorous and rapid characterization of compounds in complex matrices. The easy-to-use instrument streamlines many of the technical steps that normally require specialized setup and operation, and provides simultaneous qualitative and quantitative results. The system's Turboflow technology permits direct injection of complex matrices, virtually eliminating upfront sample preparation.

Thermo Fisher Scientific

For info: 800-532-4752 | www.thermo.com/ms

BOVINE BEADCHIP ARRAY

The next generation High-Density Bovine BeadChip array contains data from more than 20 diverse breeds generated from recent sequencing largely done with Illumina genome analyzer technology. The BovineHD can interrogate more than 500,000 loci, over 10 times more genetic markers than the manufacturer's BovineSNP50 BeadChip. The BovineHD is designed to be a powerful research tool for enabling more discoveries of quantitative traits through an expanded diversity of bovine breeds assessed in genetic prediction.

Illumina

For info: 858-882-6822 | www.illumina.com/agriculture

RESERVOIR PLATE

A new four-row partitioned reservoir plate enables aspiration of up to four reagents into a single well space, thereby saving deck space on automated liquid handling systems. The plate is partitioned into four rows with thin walls to allow pipette tips with strippers to easily reach the bottom of the channeled reservoir. Each row holds 73 ml for a total volume of 292 ml. The plates are designed for use with almost any automated liquid handling or pipetting system. The pyramid bottoms reduce reagent waste by providing a collection area with

a dead volume of less than 0.5 ml. The partitioned reservoirs are constructed of ultrapure, chemically resistant virgin polypropylene to eliminate contamination of samples or reagents with extractables and leachates from the plate material.

Porvair Sciences

For info: +44-1932-240255 | www.porvair-sciences.com

PROTEIN CHROMATOGRAPHY COLUMNS

New size-exclusion chromatography (SEC) columns for protein analysis by multiangle light scattering (MALS) are designed for resolution, reproducibility, and sensitivity. The SEC-MALS method has many applications, including quantification of protein aggregation, determination of protein conjugate stoichiometry, and confirmation of the oligomeric state of a protein. Traditional SEC columns for protein separation often shed particles, which can require down time to condition the columns. The new columns cease shedding particles after a very short time and produce exceptionally low light-scattering baseline noise and excellent pressure-shock resistance compared to other columns. This feature not only ensures high data quality and accuracy, but also improves productivity because it reduces the time waiting for the columns to "quiet down."

Wyatt Technology

For info: 805-681-9009 | www.wyatt.com

PARTICLES FOR PROTEIN ENRICHMENT

Nanotrap ESP Particles are designed to improve complex biofluid sample processing to allow for the detection of low-abundance proteins that would not otherwise be detected. The Nanotrap ESP product is designed for use with a variety of sample types that require a more efficient and powerful method of sample preparation for downstream detection and analysis. The particles improve the sensitivity of any gel electrophoresis protein detection method, including silver staining, Coomassie blue, protein immunoblotting, and mass spectrometry analysis. The particles offer the ability to harvest multiple proteins and peptides from a single sample, offering a tool for researchers interested in rapid sample processing of low-abundance proteins or for the discovery of unknown proteins and peptides in samples containing high-abundance interfering proteins.

Ceres Nanosciences

For info: 800-615-0418 | www.ceresnano.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/products/newproducts.dtl for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.